

STIC-Biotech/ChemLib

118892

From: Whiteman, Brian  
Sent: Tuesday, April 06, 2004 6:34 PM  
To: STIC-Biotech/ChemLib  
Subject: seq search

09/610,313  
7/5/00  
Barnett et al.,

search SEQ ID NO: 30, 31, and 32 against us patent and us patent application databases.

Thanks,  
Brian Whiteman  
Remsen, 2D14  
mail box 2C18  
Patent Examiner - Art Unit 1635  
United States Patent and Trademark Office  
(571) 272-0764

RECEIVED  
APR - 7 2004  
STIC

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 4/7/04  
Date Completed: 4/15/04  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: 3  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: Q66H  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: Apr-11 10, 2004, 02:53:16 ; Search time 6311.25 Seconds

(without alignments)  
16873.640 Million cell updates/sec

Title: US-09-610-313-32

Perfect score: 2457  
Sequence: 1 gtccagccaccacgcgcga.....ggactagcaccggtgaattc 2457

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_da: \*  
2: gb\_hg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_scs: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_scs: \*  
28: em\_un: \*  
29: em\_vi: \*  
30: em\_hg\_hum: \*  
31: em\_hg\_inv: \*  
32: em\_hg\_other: \*  
33: em\_hg\_mus: \*  
34: em\_hg\_pin: \*  
35: em\_hg\_rod: \*  
36: em\_hg\_mam: \*  
37: em\_hg\_vrt: \*  
38: em\_sv: \*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Match	Query Length	DB ID	Description
1	2457	100.0	2457	6	AX455916
2	2436.2	99.2	2463	6	AX455915
3	2415.4	98.3	2469	6	AX455914
4	2040	83.0	2300	6	BD263705
5	2040	83.0	2300	6	AR373388
6	2019.2	82.2	2306	6	BD263704
7	2019.2	82.2	2306	6	AR373387
8	2000.8	81.4	2312	6	AX427930
9	1998.4	81.3	2312	6	AR373389
10	1998.4	81.1	9788	6	AX427931
11	1992.4	80.5	9189	6	AX427926
12	1978.4	80.5	9189	6	AX427925
13	1975	80.3	9194	6	AX427927
14	1973.4	80.3	12411	6	AX427927
15	1973.4	80.3	9785	6	AX427938
16	1970	80.0	9157	6	AX427933
17	1966.6	80.0	9170	6	AX427928
18	1966.6	80.0	9782	6	AX427935
19	1966.6	80.0	9783	6	AX427934
20	1966.6	80.0	9792	6	AX427932
21	1966.6	80.0	9189	6	AX427921
22	1965	80.0	4353	12	AF287353
23	1962	79.9	3035	6	AX455946
24	1945.4	79.2	3009	6	AX455987
25	1928.8	78.5	4319	6	BD263639
26	1915.8	78.0	4319	6	AR373362
27	1915.8	78.0	4319	6	AR373362
28	1905.2	77.5	4352	12	AF287352
29	1878.6	76.5	1965	6	AX455952
30	1872.8	76.2	2299	6	BD263703
31	1872.8	75.7	2577	6	AX457088
32	1859.8	75.4	4343	6	AX306428
33	1852.2	75.4	2305	6	BD263702
34	1852	75.4	2305	6	AR373385
35	1852	75.2	4341	6	AX306429
36	1848.2	74.8	1977	6	AX455954
37	1838.6	74.3	7897	6	AX427923
38	1825.6	72.1	3024	12	AF287355
39	1772	72.1	9407	6	AX427937
40	1757.2	71.5	4307	6	AR302563
41	1626.6	66.2	4307	6	AX019132
42	1626.6	66.2	4307	6	BD130468
43	1626.6	65.4	4307	6	BD268901
44	1607.4	65.4	4307	6	AX035453
45	1607.4	65.4	4307	6	AX035453

ALIGNMENTS

RESULT 1  
LOCUS AX455916  
DEFINITION Sequence 32 from Patent WO0204493.  
ACCESSION AX455916  
VERSION AX455916.1 GI:21714901  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
artificial sequences.

DNA linear PAT 06-JUL-2002

CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)  
Location/Qualifiers  
1. .2457  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="PR95YMMW"

## ORIGIN

Query Match 100.0%; Score 2457; DB 6; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 2.4e-248;  
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 GTGAGCGCCACCATGGCGAGGCCATGAGCCAGGCCACCGCCACATCTCTGATGCGAG 60  
Db  
1 GTGAGCGCCACCATGGCGAGGCCATGAGCCAGGCCACCGCCACATCTCTGATGCGAG 60  
QY 61 CGCAGCACTTCAAGGGGCCCAAGCGCATCATCAAGTCTTCAACTCGGCAAGAGGGC 120  
Db 61 CGCAGCACTTCAAGGGGCCCAAGCGCATCATCAAGTCTTCAACTCGGCAAGAGGGC 120  
QY 121 CACATCGCCGCAACTGCGCGCCCGCCCGCAAGAGGGCTGCTGGAGTGGCGCAGAGG 180  
Db 121 CACATCGCCGCAACTGCGCGCCCGCCCGCAAGAGGGCTGCTGGAGTGGCGCAGAGG 180  
QY 181 GGCACACAGATGAAGGACTGCAACGAGCGCCAGGCCAACTTCTTCCGAGGACCTGGCC 240  
Db 181 GGCACACAGATGAAGGACTGCAACGAGCGCCAGGCCAACTTCTTCCGAGGACCTGGCC 240  
QY 241 TTCCCGCCAGGCAAGCGCGGAGTTCCCGAGCGAGCAGAACCGCGCCACACGCCACC 300  
Db 241 TTCCCGCCAGGCAAGCGCGGAGTTCCCGAGCGAGCAGAACCGCGCCACACGCCACC 300  
QY 301 AGCGCGAGCTCGAGTGGGGGGGAGAACCGCGCGAGCGAGGCGCGCGCGAGCGCCAG 360  
Db 301 AGCGCGAGCTCGAGTGGGGGGGAGAACCGCGCGAGCGAGGCGCGCGCGAGCGCCAG 360  
QY 361 GGCACCTTGAATTTCCCGCAGATCAACCTGTGGCAGCGCCCGCTGTGGAGTCAAGGTG 420  
Db 361 GGCACCTTGAATTTCCCGCAGATCAACCTGTGGCAGCGCCCGCTGTGGAGTCAAGGTG 420  
QY 421 GCGCGCAGATCAAGAGGCGCTGTGGACACCGCGCGCGAGCAGCACCGTGTGGAGGAG 480  
Db 421 GCGCGCAGATCAAGAGGCGCTGTGGACACCGCGCGCGAGCAGCACCGTGTGGAGGAG 480  
QY 481 ATGAGCCTGCCCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAACAAG 540  
Db 481 ATGAGCCTGCCCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAACAAG 540  
QY 541 GTGCGCAGTACGACACCATCTGTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTG 600  
Db 541 GTGCGCAGTACGACACCATCTGTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTG 600  
QY 601 CTGATCGGCGCCACCGCCGTTGAACATCATCTGCGCGCAACATGCTGACCCAGCTGGCTGC 660  
Db 601 CTGATCGGCGCCACCGCCGTTGAACATCATCTGCGCGCAACATGCTGACCCAGCTGGCTGC 660  
QY 661 ACCCTGAACTTCCCATCAGCCCATCGAGACCGTGGCGGTGAGTGAAGCCCGCGCATG 720  
Db 661 ACCCTGAACTTCCCATCAGCCCATCGAGACCGTGGCGGTGAGTGAAGCCCGCGCATG 720  
QY 721 GACGCGCCCAAGGTGAAGCACTGGCCCTGTGACCGAGGAGAGATCAAGGCCCTTGAACCGC 780  
Db 721 GACGCGCCCAAGGTGAAGCACTGGCCCTGTGACCGAGGAGAGATCAAGGCCCTTGAACCGC 780  
QY 781 ATCTGCGAGGAGATGGAGAGGGGCAAGATCAACAGATCGCCCGCGAGAACCCCTAC 840  
Db 781 ATCTGCGAGGAGATGGAGAGGGGCAAGATCAACAGATCGCCCGCGAGAACCCCTAC 840  
QY 841 AACACCCCGTGTTCGCCATCAAGAAGAGGACAGCACCAAGTGGGCGCAAGTGTGGAC 900  
Db 841 AACACCCCGTGTTCGCCATCAAGAAGAGGACAGCACCAAGTGGGCGCAAGTGTGGAC 900

QY 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCAATCCCCCAG 960  
Db 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCAATCCCCCAG 960  
QY 961 CCGCGCGCTGAAAGAAAGAGAGCGCTGACCGTCTGAGAGTGGGCGACCGCTACTTC 1020  
Db 961 CCGCGCGCTGAAAGAAAGAGAGCGCTGACCGTCTGAGAGTGGGCGACCGCTACTTC 1020  
QY 1021 AGCGTGCCTTGGACGAGGACTTCCGCAAGTACACCGCTTCAACCATCCCCAGCATCAAC 1080  
Db 1021 AGCGTGCCTTGGACGAGGACTTCCGCAAGTACACCGCTTCAACCATCCCCAGCATCAAC 1080  
QY 1081 AACGAGACCCCGGCACTCCGCTACCACTACAGTGTCTGCCCGAGGCTGGAGGGCAGC 1140  
Db 1081 AACGAGACCCCGGCACTCCGCTACCACTACAGTGTCTGCCCGAGGCTGGAGGGCAGC 1140  
QY 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGAGCCCTTCGCGCCCGCAACCCC 1200  
Db 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGAGCCCTTCGCGCCCGCAACCCC 1200  
QY 1201 GAGATCTGTATCTACCAAGCCCGCTGTACGTGGGCGAGCAGCTGGAGATCGGCAGCAG 1260  
Db 1201 GAGATCTGTATCTACCAAGCCCGCTGTACGTGGGCGAGCAGCTGGAGATCGGCAGCAG 1260  
QY 1261 GCGCCCAAGATCGAGAGCTGCGCAAGCACTCTGCTGCGTGGGGCTTCAACACCCCGCAG 1320  
Db 1261 GCGCCCAAGATCGAGAGCTGCGCAAGCACTCTGCTGCGTGGGGCTTCAACACCCCGCAG 1320  
QY 1321 AAGAAAGCAGCAGAGGAGCGCCCTTCTTCCCATCGAGTGTCAACCGGAGAGTGGAC 1380  
Db 1321 AAGAAAGCAGCAGAGGAGCGCCCTTCTTCCCATCGAGTGTCAACCGGAGAGTGGAC 1380  
QY 1381 GTGAGCCCATCGAGCTGCGGAGAGAGAGTGAACCGTGAACGACATCCAGAGCTG 1440  
Db 1381 GTGAGCCCATCGAGCTGCGGAGAGAGAGTGAACCGTGAACGACATCCAGAGCTG 1440  
QY 1441 GTGGGCAAGCTGAACCTGGGCGACCCAGATCTACCCCGCATCAAGTGTGGCGCTGTGC 1500  
Db 1441 GTGGGCAAGCTGAACCTGGGCGACCCAGATCTACCCCGCATCAAGTGTGGCGCTGTGC 1500  
QY 1501 AAGTGTGCGCGCGCCAAAGGCCCTGACCGACATGCTGCCCTGACCGAGAGCGCGAG 1560  
Db 1501 AAGTGTGCGCGCGCCAAAGGCCCTGACCGACATGCTGCCCTGACCGAGAGCGCGAG 1560  
QY 1561 CTGAGCTGGCGGAGAACCGCGAGATCTTGGCGAGCGCGCTGACCGAGCTGTACTACGAC 1620  
Db 1561 CTGAGCTGGCGGAGAACCGCGAGATCTTGGCGAGCGCGCTGACCGAGCTGTACTACGAC 1620  
QY 1621 CCCAGCAAGCACTGTGGCGGAGATCCAGAGCGAGGCGCACGACAGTGGACCTTACCAG 1680  
Db 1621 CCCAGCAAGCACTGTGGCGGAGATCCAGAGCGAGGCGCACGACAGTGGACCTTACCAG 1680  
QY 1681 ATCTACCGAGAGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGGCACCGCC 1740  
Db 1681 ATCTACCGAGAGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGGCACCGCC 1740  
QY 1741 CACACCAAGCTGTGAAGCAGCTGACCGAGGCGCTGCAAGATCGCATGAGAGGATC 1800  
Db 1741 CACACCAAGCTGTGAAGCAGCTGACCGAGGCGCTGCAAGATCGCATGAGAGGATC 1800  
QY 1801 GTGATCTGGGCGCAAGCCCGCAAGTTCGCTTCCCGCTGCGCATCCAGAGGAGACCTGGGAGAC 1860  
Db 1801 GTGATCTGGGCGCAAGCCCGCAAGTTCGCTTCCCGCTGCGCATCCAGAGGAGACCTGGGAGAC 1860  
QY 1861 TGGTGAACCGACTTACTGGCAGCGCCACTGGATTCCTCGAGTGGAGTTCGTGAACACCC 1920  
Db 1861 TGGTGAACCGACTTACTGGCAGCGCCACTGGATTCCTCGAGTGGAGTTCGTGAACACCC 1920  
QY 1921 CCGCTGTGAAGCTGTGGTACCAGCTGGAGAGAGGAGCCCATCATCGGCGCGGAGACCTTC 1980  
Db 1921 CCGCTGTGAAGCTGTGGTACCAGCTGGAGAGAGGAGCCCATCATCGGCGCGGAGACCTTC 1980  
QY 1981 TACGTGAGCGCGCCCGCCAAACCGCGGAGACCAAGATCGGCAAGGCGCGCTACGTGACCGAC 2040

1981 TACGTGGACGGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTGACCGAC 2040  
2041 CGGGCGCGCAGAGATCGTGGACCTGACCGAGACCAACCAAGAGACCGAGTGCAG 2100  
2041 CGGGCGCGCAGAGATCGTGGACCTGACCGAGACCAACCAAGAGACCGAGTGCAG 2100  
2101 GCCATCCAGTGGCCCTGACGACACCGCGACCGAGGTGAACATCGTACCGACAGCGAG 2160  
2101 GCCATCCAGTGGCCCTGACGACACCGCGACCGAGGTGAACATCGTACCGACAGCGAG 2160  
2161 TAGCCCTGGCATCTACCGCGCCAGCGCGACCGAGCGAGCGAGTGTGAACAG 2220  
2161 TAGCCCTGGCATCTACCGCGCCAGCGCGACCGAGCGAGCGAGTGTGAACAG 2220  
2221 ATCATCGAGCAGTGTATCAAGAGAGAGAGTGTACTTACCTGAGTGGTGGCCGCCACAG 2280  
2221 ATCATCGAGCAGTGTATCAAGAGAGAGAGTGTACTTACCTGAGTGGTGGCCGCCACAG 2280  
2281 GGCATCGCGCGCAACGACGACATCGACAGTGTGTAGCAAGGGCATCGCAGAGTGTG 2340  
2281 GGCATCGCGCGCAACGACGACATCGACAGTGTGTAGCAAGGGCATCGCAGAGTGTG 2340  
2341 TTCTCTGGACGGCATCGATGGCGCATCGTGTATCTACAGTACATGACGACCTGTACGTG 2400  
2341 TTCTCTGGACGGCATCGATGGCGCATCGTGTATCTACAGTACATGACGACCTGTACGTG 2400  
2401 GGCAGCGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457  
2401 GGCAGCGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457

RESULT 2  
AX455915  
LOCUS AX455915 2463 bp DNA linear PAT 06-JUL-2002  
DEFINITION Sequence 31 from Patent WO0204493.  
ACCESSION AX455915  
VERSION AX455915.1 GI:21714900

KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
artificial sequences.

REFERENCE  
1. zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.

AUTHORS Polynucleotides encoding antigenic hiv type c polypeptides,  
TITLE polypeptides and uses thereof

JOURNAL Patent: WO 0204493-A 31 17-JAN-2002;

CHIRON CORPORATION (US); University of Stellenbosch (ZA)

FEATURES  
Location/Qualifiers

1. 2463  
/organism="synthetic construct"  
/mol type="unassigned DNA"  
/db xref="taxon:32630"  
/note="PR975YN"

## ORIGIN

Query Match 99.2%; Score 2436.2; DB 6; Length 2463;  
Best Local Similarity 99.6%; Pred. No. 3.6e-246;  
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;  
1 GTCAGCGCCACCATGGCGGAGGCGCATGAGCGAGCGCCACCGCGCCACCATCTGTGTCAG 60  
1 GTCAGCGCCACCATGGCGGAGGCGCATGAGCGAGCGCCACCGCGCCACCATCTGTGTCAG 60  
61 GCGAGCAACTTCAAGGGCGCCCGACGCGATCATCAAGTGTCTCACTCGCGCGAGAGGGC 120  
61 GCGAGCAACTTCAAGGGCGCCCGACGCGATCATCAAGTGTCTCACTCGCGCGAGAGGGC 120  
121 CACATCGCGCGCAACTGCGCGCGCCCGCGAGAGGGGTGCTGGAAGTGGCGCAAGGAG 180  
121 CACATCGCGCGCAACTGCGCGCGCCCGCGAGAGGGGTGCTGGAAGTGGCGCAAGGAG 180

181 GGCACACAGATGAAGACTGACACCGAGCGCAGGCCAATTCTTCCGCGAGGACTGGCC 240  
241 TTCCCCCAGAGCGCCCGCGAGTTCCCCAGCGAGCAAGAACCGCCCAACAGCCCAACC 300  
241 TTCCCCCAGAGCGCCCGCGAGTTCCCCAGCGAGCAAGAACCGCCCAACAGCCCAACC 300  
301 AGCGCGAGCTGAGAGTGGCGCGGAGAACCCCGCGAGCGAGCGCGCGCGAGCGCCAG 360  
301 AGCGCGAGCTGAGAGTGGCGCGGAGAACCCCGCGAGCGAGCGCGCGCGAGCGCCAG 360  
361 GGCACCTGAACTTCCCGCAGATCACCTGTGCGAGCGCCCTGTGTGAGCATCAAGGTG 420  
361 GGCACCTGAACTTCCCGCAGATCACCTGTGCGAGCGCCCTGTGTGAGCATCAAGGTG 420  
421 GCGCGCAGATCAAGAGAGCGCTGTGTGACACCGCGCGCGAGCACCGGTGTGGAGGAG 480  
421 GCGCGCAGATCAAGAGAGCGCTGTGTGACACCGCGCGCGCGAGCACCGGTGTGGAGGAG 480  
481 ATGAGCTGCGCGCAAGTGAAGCGCCCAAGATCATCGCGCGCATCGCGCGCTTCATCAAG 540  
481 ATGAGCTGCGCGCAAGTGAAGCGCCCAAGATCATCGCGCGCATCGCGCGCTTCATCAAG 540  
541 GTGCGCAGTACGACAGATCTGTGATCGAGATCTGCGCGCAAGAGCGCATCGGACCGTG 600  
541 GTGCGCAGTACGACAGATCTGTGATCGAGATCTGCGCGCAAGAGCGCATCGGACCGTG 600  
601 CTGATCGCGCGCCACCGCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGC 660  
601 CTGATCGCGCGCCACCGCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGC 660  
661 ACCCTGAACTTCCCGCATCAGCCCCATCGAGACCGTGCCTGTAAGCTGAAGCGCGCATG 720  
661 ACCCTGAACTTCCCGCATCAGCCCCATCGAGACCGTGCCTGTAAGCTGAAGCGCGCATG 720  
721 GACGCGCGCGCGCGCGCGTGAAGCGCGTGCCTGACCGAGAGAGATCAAGCGCCCTCACGCC 780  
721 GACGCGCGCGCGCGCGCGTGAAGCGCGTGCCTGACCGAGAGAGATCAAGCGCCCTCACGCC 780  
781 ATCTGCGAGGATGGAGAGAGGCGCAAGATCAACAGATCGCGCGCGCGCGCGCATAC 840  
781 ATCTGCGAGGATGGAGAGAGGCGCAAGATCAACAGATCGCGCGCGCGCGCGCATAC 840  
841 AACACCGCGCGTTCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
841 AACACCGCGCGTTCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
901 TTCGCGAGCTGAACAAGCGCACCGAGACTTCTGGAGAGTGCAGCTGGGCGATCCCCCAC 960  
901 TTCGCGAGCTGAACAAGCGCACCGAGACTTCTGGAGAGTGCAGCTGGGCGATCCCCCAC 960  
961 CCCGCGCGCGCTGAAG 1020  
961 CCCGCGCGCGCTGAAG 1020  
1021 AGCGTGCCTGCGAGCGAG 1080  
1021 AGCGTGCCTGCGAGCGAG 1080  
1081 AACAG 1140  
1081 AACAG 1140  
1141 CCCAG 1200  
1141 CCCAG 1200  
1201 GAGATCGTGTATTAACAGCG 1260  
1201 GAGATCGTGTATTAACAGCG 1260



Db 1261 CGCGCAAGATCGAGGAGCTCGCAAGCACTGCTGGCTGGGCTTCAACACCCCGCAC 1320  
 QY 1321 AAGAGCACCAAGAGGAGCCCTCTCTGCCAT-----CGAGCTGCAACCCGCAAG 1374  
 Db 1321 AAGAGCACCAAGAGGAGCCCTCTCTGGATGGGTACGAGCTGCAACCCGCAAG 1380  
 QY 1375 TGGACCGTGCAGCCATCGACTCCCGAGAGAGAGCTGGACCGTGAACGACATCCAG 1434  
 Db 1381 TGGACCGTGCAGCCATCGACTCCCGAGAGAGAGCTGGACCGTGAACGACATCCAG 1440  
 QY 1435 AAGCTGTGGGCAAGCTGAAGTGGGCGAGCCAGATCTACCCCGCATCAAGGTGGCCAG 1494  
 Db 1441 AAGCTGTGGGCAAGCTGAAGTGGGCGAGCCAGATCTACCCCGCATCAAGGTGGCCAG 1500  
 QY 1495 CTGTGCAAGCTGCTGCGCGGCGCAAGCCCTGACCGACATCGTGTGCCCTGACCGAGGAG 1554  
 Db 1501 CTGTGCAAGCTGCTGCGCGGCGCAAGCCCTGACCGACATCGTGTGCCCTGACCGAGGAG 1560  
 QY 1555 GCGAGCTGGAGCTGGCGAGAACCGGAGATCTGCGGAGCCGTCGAGCGGTGTAC 1614  
 Db 1561 GCGAGCTGGAGCTGGCGAGAACCGGAGATCTGCGGAGCCGTCGAGCGGTGTAC 1620  
 QY 1615 TACGACCCAGCAAGGACCTGTGTGCGGAGATCCAGAGCAGGCGCCACGACGTGAC 1674  
 Db 1621 TACGACCCAGCAAGGACCTGTGTGCGGAGATCCAGAGCAGGCGCCACGACGTGAC 1680  
 QY 1675 TACGAGTCTACGAGGACCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATCGC 1734  
 Db 1681 TACGAGTCTACGAGGACCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATCGC 1740  
 QY 1735 ACCGCGCACCAAGCAGCTGAAGCAGCTGACCGAGGCGTGCAAGATCGCCATCGAG 1794  
 Db 1741 ACCGCGCACCAAGCAGCTGAAGCAGCTGACCGAGGCGTGCAAGATCGCCATCGAG 1800  
 QY 1795 AGCATCTGTGATCTGGGCAAGACCCCAAGTTCCGCTTGGCTTCCATCCAGAGGACCTGG 1854  
 Db 1801 AGCATCTGTGATCTGGGCAAGACCCCAAGTTCCGCTTGGCTTCCATCCAGAGGACCTGG 1860  
 QY 1855 GAGACCTGTGAGCAGCTACTTGGCAGGCGCACTGTGATCCCGGAGTGGAGTTCTGTAA 1914  
 Db 1861 GAGACCTGTGAGCAGCTACTTGGCAGGCGCACTGTGATCCCGGAGTGGAGTTCTGTAA 1920  
 QY 1915 ACCCCCCCTGTGTGAAGCTGTGTACCAAGTGTGAGAGAGGAGCCATCATCGGCGCGAG 1974  
 Db 1921 ACCCCCCCTGTGTGAAGCTGTGTACCAAGTGTGAGAGAGGAGCCATCATCGGCGCGAG 1980  
 QY 1975 ACCTTCTAGTGGAGCGGCGCCCAACCGGAGACCAAGATCGGAGGCGGCTAGCTG 2034  
 Db 1981 ACCTTCTAGTGGAGCGGCGCCCAACCGGAGACCAAGATCGGAGGCGGCTAGCTG 2040  
 QY 2035 ACCGACCGGCGCGCAGAAGATCTGTGAGCTGTGACCGAGACCAACCAAGAGACCGAG 2094  
 Db 2041 ACCGACCGGCGCGCAGAAGATCTGTGAGCTGTGACCGAGACCAACCAAGAGACCGAG 2100  
 QY 2095 CTGACGCGCATCAGCTGGCTGTGAGAGCAGCAGCGAGGAGTGAACATCGTACCGAC 2154  
 Db 2101 CTGACGCGCATCAGCTGGCTGTGAGAGCAGCAGCGAGGAGTGAACATCGTACCGAC 2160  
 QY 2155 AGCCAGTACGCTTGGGATCATCCAGGCGCCAGCCCGCAAGAGGAGAGGAGCTGGT 2214  
 Db 2161 AGCCAGTACGCTTGGGATCATCCAGGCGCCAGCCCGCAAGAGGAGAGGAGCTGGT 2220  
 QY 2215 AACGAGTACGAGCAGCTGATCAAGAGGAGAGGAGTGTACCTGAGTGGTGGTCCGCC 2274  
 Db 2221 AACGAGTACGAGCAGCTGATCAAGAGGAGAGGAGTGTACCTGAGTGGTGGTCCGCC 2280  
 QY 2275 CACAAGGAGCATCGGCGGCAACGAGCAGATCGCAAGCTGGTGAAGAGGAGATCCGCAAG 2334  
 Db 2281 CACAAGGAGCATCGGCGGCAACGAGCAGATCGCAAGCTGGTGAAGAGGAGATCCGCAAG 2340  
 QY 2335 GTGCTGTCTGAGCGGCAATCGATGGCGCATCTGTGATCTACCACTGATGAGACACCTG 2394  
 Db 2341 GTGCTGTCTGAGCGGCAATCGATGGCGCATCTGTGATCTACCACTGATGAGACACCTG 2400

QY 2395 TACGTGGCAGCGCGCCCTAGGATGATTAAAGCTTCCCGGGTAGCACCGGTGAA 2454  
 Db 2401 TACGTGGCAGCGCGCCCTAGGATGATTAAAGCTTCCCGGGTAGCACCGGTGAA 2460  
 QY 2455 TTC 2457  
 Db 2461 TTC 2463  
 RESULT 3  
 AX455914  
 LOCUS AX455914 2469 bp DNA linear PAT 06-JUL-2002  
 DEFINITION Sequence 30 from Patent WO2004493.  
 ACCESSION AX455914  
 VERSION AX455914.1 GI:21714899  
 KEYWORDS synthetic construct  
 SOURCE synthetic construct  
 ORGANISM artificial sequences.  
 1  
 REFERENCE zur Mesede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.  
 AUTHORS Polynucleotides encoding antigenic hiv type c polypeptides,  
 TITLE polypeptides and uses thereof  
 JOURNAL Patent: WO 0204493-A 30 17-JAN-2002;  
 CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)  
 FEATURES  
 Location/Qualifiers  
 1..2469  
 /organism="synthetic construct"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32630"  
 /note="PR975(+)"  
 ORIGIN  
 Query Match 98.3%; Score 2415.4; DB 6; Length 2469;  
 Best Local Similarity 99.3%; Pred. No. 5.3e-244; Indels 12; Gaps 2;  
 Matches 2451; Conservative 0; Mismatches 6;  
 QY 1 GTGACGCCACCATGCGCGAGCCCATGAGCCAGGCCACAGCGCCAAATCTGTATGAG 60  
 Db 1 GTGACGCCACCATGCGCGAGCCCATGAGCCAGGCCACAGCGCCAAATCTGTATGAG 60  
 QY 61 CGCAGCACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCACTGGGCAAGGAGGC 120  
 Db 61 CGCAGCACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCACTGGGCAAGGAGGC 120  
 QY 121 CACATCGCCGCAACTGCGCGCCCGCCCGCCCGCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGGAG 180  
 Db 121 CACATCGCCGCAACTGCGCGCCCGCCCGCCCGCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGGAG 180  
 QY 181 GGCACACAGATGAAGACTGTGACCGCGCCAGCGCCAGGCGCAATCTTTCGCGAGGACCTGGCC 240  
 Db 181 GGCACACAGATGAAGACTGTGACCGCGCCAGCGCCAGGCGCAATCTTTCGCGAGGACCTGGCC 240  
 QY 241 TTCCCGCAGGCAAGGCCCGCGAGTTCGCCAGCGAGAACCCGCGCCAAACCGCCCAACCGCCACC 300  
 Db 241 TTCCCGCAGGCAAGGCCCGCGAGTTCGCCAGCGAGAACCCGCGCCAAACCGCCCAACCGCCACC 300  
 QY 301 AGCCGCGAGCTGCGAGTGTGCGCGCGAGAACCCCGCGCGAGGCGCGCGCGCGAGCGCCAG 360  
 Db 301 AGCCGCGAGCTGCGAGTGTGCGCGCGAGAACCCCGCGCGAGGCGCGCGCGCGAGCGCCAG 360  
 QY 361 GGCACCTGAACCTTCCCGCAGATCACCTGTGGCAGCCCTTGTGTGAGCATCAAGGTG 420  
 Db 361 GGCACCTGAACCTTCCCGCAGATCACCTGTGGCAGCCCTTGTGTGAGCATCAAGGTG 420  
 QY 421 GCGCGCCAGATCAAGGAGGCCCTGTGTGACACCGCGCCGACACACCGCTGTGTGAGGAG 480  
 Db 421 GCGCGCCAGATCAAGGAGGCCCTGTGTGACACCGCGCCGACACACCGCTGTGTGAGGAG 480  
 QY 481 ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTCAATCAAG 540  
 Db 481 ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTCAATCAAG 540

541 GTGGCCAGTACGACCATCTGATCGAGATCTCGGCAAGAAAGGCCATCGGACCGTG 600  
Db  
541 GTGGCCAGTACGACCATCTGATCGAGATCTCGGCAAGAAAGGCCATCGGACCGTG 600  
Qy  
601 CTGATCGGCCACCCCGTGAACATCATCGGCGCAACATCTGACCCAGCTGGGTGC 660  
Db  
601 CTGATCGGCCACCCCGTGAACATCATCGGCGCAACATCTGACCCAGCTGGGTGC 660  
Qy  
661 ACCCTGAATCTCCCATCAGCCCATCGAGACCGTGCCTGAAAGCTGAAGCCCGGCATG 720  
Db  
661 ACCCTGAATCTCCCATCAGCCCATCGAGACCGTGCCTGAAAGCTGAAGCCCGGCATG 720  
Qy  
721 GACGCGCCCAAGGTGAAGAGTGCGCCCTGACCGAGGAGAAATCAAGGCCCTGACCGCC 780  
Db  
721 GACGCGCCCAAGGTGAAGAGTGCGCCCTGACCGAGGAGAAATCAAGGCCCTGACCGCC 780  
Qy  
781 ATCTCGAGGAGATCGAGAGGAGGCAAGATCAACAAAGTCCGGCCCGAGAAACCCCTAC 840  
Db  
781 ATCTCGAGGAGATCGAGAGGAGGCAAGATCAACAAAGTCCGGCCCGAGAAACCCCTAC 840  
Qy  
841 AACACCCCGCTTTCGCCATCAAGAAAGAGAGACACCAAGTGGCGCAAGCTGGTGAC 900  
Db  
841 AACACCCCGCTTTCGCCATCAAGAAAGAGAGACACCAAGTGGCGCAAGCTGGTGAC 900  
Qy  
901 TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTCGGAGGTGCAGCTGGGCATCCCCAC 960  
Db  
901 TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTCGGAGGTGCAGCTGGGCATCCCCAC 960  
Qy  
961 CCCGCGCGCTGAAGAAAGAGAGCGTGAACCGTGTGAGCGTGGCGAGCGCTACTTC 1020  
Db  
961 CCCGCGCGCTGAAGAAAGAGAGCGTGAACCGTGTGAGCGTGGCGAGCGCTACTTC 1020  
Qy  
1021 AGCGTCCCTCGAGAGACTTTCGCAAGTACACCGCTTCAACCATCCCGAGCATCAAC 1080  
Db  
1021 AGCGTCCCTCGAGAGACTTTCGCAAGTACACCGCTTCAACCATCCCGAGCATCAAC 1080  
Qy  
1081 AACAGAGACCCCGGATCCGCTACCAAGTACACGCTGCTGCCCGAGGCTGGAAGGCGAGC 1140  
Db  
1081 AACAGAGACCCCGGATCCGCTACCAAGTACACGCTGCTGCCCGAGGCTGGAAGGCGAGC 1140  
Qy  
1141 CCCAGATCTTCAGAGAGCATGACCAAGATCTCGAGCGCTTCCGCGCGCGCAACCCC 1200  
Db  
1141 CCCAGATCTTCAGAGAGCATGACCAAGATCTCGAGCGCTTCCGCGCGCGCAACCCC 1200  
Qy  
1201 GAGATCGTATCTACCA-----GGCCCCCTGTACGTGGGCGAGCGCTGAGATCGGC 1254  
Db  
1201 GAGATCGTATCTACCAAGTACATGACGACGCTGTACGTGGGCGAGCGCTGAGATCGGC 1254  
Qy  
1255 CAGCACCGCGCCCAAGATCGAGAGCTGCGCAAGCACTGTGCTGCGTGGGGCTTCAACACC 1314  
Db  
1261 CAGCACCGCGCCCAAGATCGAGAGCTGCGCAAGCACTGTGCTGCGTGGGGCTTCAACACC 1320  
Qy  
1315 CCCGCAAGAGACCAAGAGAGCGCCCTTCTCTGCCCCAT-----CGAGTCGACCCC 1368  
Db  
1321 CCCGCAAGAGACCAAGAGAGCGCCCTTCTCTGCCCCATGCGTGGGCTACGAGTCGACCCC 1380  
Qy  
1369 GACAGTGCACGTCAGCCCATCGAGCTGCCCGAGAGAGAGCTGAGCCGTGACGAC 1428  
Db  
1381 GACAGTGCACGTCAGCCCATCGAGCTGCCCGAGAGAGAGCTGAGCCGTGAGCCGAC 1440  
Qy  
1429 ATCCAGAGCTGGTGGGCAAGCTGAACCTGGGCGAGCGATCTACCCCGGCATCAAGGTG 1488  
Db  
1441 ATCCAGAGCTGGTGGGCAAGCTGAACCTGGGCGAGCGATCTACCCCGGCATCAAGGTG 1500  
Qy  
1489 CGCAGCTGTCAAGCTGTGCGCGCGCCCAAGGCCCTGACCGCATGTGCGCCCTGACC 1548  
Db  
1501 CGCAGCTGTCAAGCTGTGCGCGCGCCCAAGGCCCTGACCGCATGTGCGCCCTGACC 1560  
Qy  
1549 GAGGAGCGAGCTGAGCTGCGCGAGAGACCGGAGATCTCTCGGAGCGCGCTGACGCGC 1608  
Db  
1561 GAGGAGCGAGCTGAGCTGCGCGAGAGACCGGAGATCTCTCGGAGCGCGCTGACGCGC 1620

Qy  
1609 GTGTACTACGACCCCGAGCAAGGACCTGTGTGGCCGAGATCCAGAAACAGGGCCACACGAC 1668  
Db  
1621 GTGTACTACGACCCCGAGCAAGGACCTGTGTGGCCGAGATCCAGAAACAGGGCCACACGAC 1680  
Qy  
1669 TGGACCTACAGATCTTACCAGAGCCCTTCAAGAAACCTTGAAGACCGGCAAGTACGCCAAG 1728  
Db  
1681 TGGACCTACAGATCTTACCAGAGCCCTTCAAGAAACCTTGAAGACCGGCAAGTACGCCAAG 1740  
Qy  
1729 ATGCGCACCGCCACACCAACGAGCGTGAAGAGTGAACGAGCGCGTGCAGAAAGATCGCC 1788  
Db  
1741 ATGCGCACCGCCACACCAACGAGCGTGAAGAGTGAACGAGCGCGTGCAGAAAGATCGCC 1800  
Qy  
1789 ATGAGAGAGATCTGTATCTGGGCGCAAGACCCCAAGTTCGCTCCCATCCAGAAAGAG 1848  
Db  
1801 ATGAGAGAGATCTGTATCTGGGCGCAAGACCCCAAGTTCGCTCCCATCCAGAAAGAG 1860  
Qy  
1849 ACTTGGGAGACCTGTGTGAACCGACTACTGTGGCGGCAACCTGTATCCCGAGTGGGAGTTC 1908  
Db  
1861 ACTTGGGAGACCTGTGTGAACCGACTACTGTGGCGGCAACCTGTATCCCGAGTGGGAGTTC 1920  
Qy  
1909 GTGAAACACCCCGCTGTGTGAAGCTGTGTACCGAGTGTGAGAGAGAGAGAGAGAGAGAG 1968  
Db  
1921 GTGAAACACCCCGCTGTGTGAAGCTGTGTACCGAGTGTGAGAGAGAGAGAGAGAGAGAG 1980  
Qy  
1969 GCCGAGACCTTCTACGTGAGACCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGCGC 2028  
Db  
1981 GCCGAGACCTTCTACGTGAGACCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGCGC 2040  
Qy  
2029 TAGGTACCGAGCGCGCGCGCGAGAGATCTGTGAGCTGTGAGAGAGAGAGAGAGAGAGAG 2088  
Db  
2041 TAGGTACCGAGCGCGCGCGCGAGAGATCTGTGAGCTGTGAGAGAGAGAGAGAGAGAGAG 2100  
Qy  
2089 ACCGAGCTGAGGCGCATCCAGCTGTGGCCCTGTGAGGACAGCGGCGAGCGAGGTGAACATCGTG 2148  
Db  
2101 ACCGAGCTGAGGCGCATCCAGCTGTGGCCCTGTGAGGACAGCGGCGAGCGAGGTGAACATCGTG 2160  
Qy  
2149 ACCGAGCGAGTACCGCTTGGCGATCATCGAGCGCGCGCGCGCGAGAGAGAGAGAGAGAG 2208  
Db  
2161 ACCGAGCGAGTACCGCTTGGCGATCATCGAGCGCGCGCGCGCGAGAGAGAGAGAGAGAG 2220  
Qy  
2209 CTGCTGAACAGAGATCATCGAGCGAGTGTATCAAGAGAGAGAGTGTATCTGAGCTGGGTG 2268  
Db  
2221 CTGCTGAACAGAGATCATCGAGCGAGTGTATCAAGAGAGAGAGTGTATCTGAGCTGGGTG 2280  
Qy  
2269 CCCGCCCAAGGCGCATCGCGCGCAACAGAGAGATCGAACAGCTGTGTGAGAGAGGCGATC 2328  
Db  
2281 CCCGCCCAAGGCGCATCGCGCGCAACAGAGAGATCGAACAGCTGTGTGAGAGAGGCGATC 2340  
Qy  
2329 CGCAAGTGTGTCTGTGAGCGGCGATCGATGGCGCGATCGTATCTACCAAGTACATGGAC 2388  
Db  
2341 CGCAAGTGTGTCTGTGAGCGGCGATCGATGGCGCGATCGTATCTACCAAGTACATGGAC 2400  
Qy  
2389 GACCTGTACGTGGCGAGCGCGCGCTTAGGATCGATTAAAGCTTCCCGGGGTAGCAC 2448  
Db  
2401 GACCTGTACGTGGCGAGCGCGCGCTTAGGATCGATTAAAGCTTCCCGGGGTAGCAC 2460  
Qy  
2449 GGTCAATTC 2457  
Db  
2461 GGTCAATTC 2469

## RESULT 4

BD263705  
LOCUS  
DEFINITION Improved expression of HIV polypeptides and production of virus-like particles.  
ACCESSION BD263705  
VERSION BD263705.1  
KEYWORDS GI:33073473  
SOURCE JP 200253124-A/72.  
ORGANISM synthetic construct  
REFERENCE 1 (bases 1 to 2300).  
PAT 17-JUL-2003

AUTHORS Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.  
 TITLE Improved expression of HIV polypeptides and production of virus-like particles  
 JOURNAL Patent: JP 2002533124-A 72 08-OCT-2002;  
 COMMENT CHIRON CORP  
 OS Artificial Sequence  
 PN JP 2002533124-A/72  
 PD 30-DEC-1999  
 PF 30-DEC-1999 JP 2000591193  
 PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 PI  
 SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI  
 KARIN HARTOG,  
 PI HONG LIU,KATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC  
 C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC  
 C12N5/10,  
 PC C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC  
 Description of Artificial Sequence: FS(-).promod.Rtopt.YMMH FH  
 KEY Location/Qualifiers  
 FT source 1..2300  
 /organism='Artificial Sequence'.  
 /location/Qualifiers  
 1..2300  
 /organism='synthetic construct'  
 /mol\_type='genomic DNA'  
 /db\_xref='taxon:32630'  
 ORIGIN  
 Query Match 83.0%; Score 2040; DB 6; Length 2300;  
 Best Local Similarity 93.6%; Pred. No. 1e-204;  
 Matches 2153; Conservative 0; Mismatches 135; Indels 12; Gaps 2;  
 QY 170 GCGGCAAGGCGCCACAGATGAAGACTGCACCGAGCGCCAGCCCAACTTCTTCGCG 229  
 DB 1 GCGGCGGCAAGGACCAAAATGAAGATTGCATCGAGAGACAGGCTTAATTCTTCGCG 60  
 QY 230 AGGACTCGCTTCCCGCAGGCGAAGGCGCGAGTTCGCCAGGAGCAGAACCCGCCA 289  
 DB 61 AGGACTCGCTTCTCTGAGGCGAAGGCGCGAGTTCAGCAGCGAGCAGCCGCCA 120  
 QY 290 ACAGCCCAACAGCGCGAGTGCAGGTGCGCGCG-----ACACCCCGCAGCGAGG 343  
 DB 121 ACAGCCCAACCGCGCGAGTGCAGGTGCGCGCGGAGAGCAACAGCGCTGAGCGAGG 180  
 QY 344 CCGGCGCGAGCGCGCAGCGACCCCTG-----AACTTCCCGCAGATCACTCTGGCAGC 397  
 DB 181 CCGGCGCGAGCGCGCAGCGACCGTGAAGTTCACCTTCCCGCAGATCACTCTGGCAGC 240  
 QY 398 GCCCGCTGGTGAAGCATCAAGGTGGCGCGCGAGATCAAGAGGCGCTTGTGAGACCGGCG 457  
 DB 241 GCCCGCTGGTGAAGCATCAAGGTGGCGCGCGAGTCAAGAGGCGCGCTGTGAGACCGGCG 300  
 QY 458 CCGAGCACACCTGTGGAGAGATGAGCTGCCCGCAAGTGGAGGCCCAAGATGATCG 517  
 DB 301 CCGAGCACACCTGTGGAGAGATGAGCTGCCCGCAAGTGGAGGCCCAAGATGATCG 360  
 QY 518 GCGGATCGCGGCTTCATCAAGGTGGCGCGAGTACGACACAGATCTGTGAGATCTGCG 577  
 DB 361 GCGGATCGCGGCTTCATCAAGGTGGCGCGAGTACGACACAGATCTGTGAGATCTGCG 420  
 QY 578 GCAAGAAGCCATCGGACCGTGTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 637  
 DB 421 GCCACAGGCCATCGGACCGTGTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 480  
 QY 638 ACATCTGACCGAGTGGCTGCACCTTGAATCTTCCCATCAGCCCGCATCGAGCCGTCG 697  
 DB 481 ACCTGTGACCGAGATCGGCTGCACCTTGAATCTTCCCATCAGCCCGCATCGAGCCGTCG 540  
 QY 698 CCGTGAAGTGAAGCCCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 757  
 DB 541 CCGTGAAGTGAAGCCCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
 QY 758 AGAAGATCAAGCCCTGACCGCGCATCTGCGAGGAGATGGAGAGGAGGCGCAAGATCACCA 817

DB 601 AGAAGATCAAGGCGCTGTGTGAGATCTGCACCGAGATGGAGAGGAGCGCAAGATCAGCA 660  
 QY 818 AGATCGGCGCGGAGAACCCCTTACACACCCCGGTGTGCGCATCAAGAAGAGCAGCA 877  
 DB 661 AGATCGGCGCGGAGAACCCCTTACACACCCCGGTGTGCGCATCAAGAAGAGCAGCA 720  
 QY 878 CCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAACAAGCGACCCAGAGCTTCTGGG 937  
 DB 721 CCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAACAAGCGACCCAGAGCTTCTGGG 780  
 QY 938 AGTGTGAGTGGGATCCCGACCCCGCGCGCTTGAAGAAGAGAGAGCTGACCGTGC 997  
 DB 781 AGTGTGAGTGGGATCCCGACCCCGCGCGCTTGAAGAAGAGAGAGCTGACCGTGC 840  
 QY 998 TGGAGTGGGCGACCGCTTACTTACGCTGCGCCCTGGAGAGGACTTCCGCAAGTACACCG 1057  
 DB 841 TGGAGTGGGCGACCGCTTACTTACGCTGCGCCCTGGAGAGGACTTCCGCAAGTACACCG 900  
 QY 1058 CTTTACACATCCCGACGATCAACAAGAGAGACCCCGCGATCGCTACAGTACACGTCG 1117  
 DB 901 CTTTACACATCCCGACGATCAACAAGAGAGACCCCGCGATCGCTACAGTACACGTCG 960  
 QY 1118 TGCCCCAGGCTGGAAGGCGACCCCGCATCTTCCAGAGCAGATGACCAAGATCTCTGG 1177  
 DB 961 TGCCCCAGGCTGGAAGGCGACCCCGCATCTTCCAGAGCAGATGACCAAGATCTCTGG 1020  
 QY 1178 AGCCCTTCCGCGCGCAACCCCGAGATCGTGTATCTACAGCCCGCTGTACGTGGCA 1237  
 DB 1021 AGCCCTTCCGCGCAAGAGAACCCCGCATCTTACAGGCGCCCGCTGTACGTGGCA 1080  
 QY 1238 GCGACCTGAGATCGGCGAGCAGCCCGCAAGATCGAGAGCTGGCGCAAGCAGCTGTCTGC 1297  
 DB 1081 GCGACCTGAGATCGGCGAGCAGCCCGCAAGATCGAGAGCTGGCGCAAGCAGCTGTCTGC 1140  
 QY 1298 GTTGGGCTTACACCCCGCAAGAAGCAGCAGAGAGCGCCCTTCTGCGCCATCG 1357  
 DB 1141 GTTGGGCTTACACCCCGCAAGAAGCAGCAGAGAGCGCCCTTCTGCGCCATCG 1200  
 QY 1358 AGCTGACCCCGCAGAGTGGAGCCGTGCGAGCCCATCGAGCTGCCAGAGAGAGAGCTGA 1417  
 DB 1201 AGCTGACCCCGCAGAGTGGAGCCGTGCGAGCCCATCTGCTGCCCGAGAGAGCAGCTGA 1260  
 QY 1418 CCGTGAACGACATCCAGAAAGCTGGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCG 1477  
 DB 1261 CCGTGAACGACATCCAGAAAGCTGGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCG 1320  
 QY 1478 GCATCAAGTGGCGCAGCTGTGCAAGCTGTGCGGCGCGCAAGCCCTGACCGCATCG 1537  
 DB 1321 GCATCAAGTGAAGCAGCTGTGCAAGCTGTGCGGCGCAGCAGCCCTGACCGAGTGA 1380  
 QY 1538 TGCCCCCTGACCGAGGAGCGGAGCTGGAGCTGGCGAGAACCGCGAGATCTTGGCGGAGC 1597  
 DB 1381 TGCCCCCTGACCGAGGAGCGGAGCTGGAGCTGGCGAGAACCGCGAGATCTTGAAGAGAGC 1440  
 QY 1598 CCGTGAAGCTGTGATCTACGAGCCCGCAGAGCCTGTGGCGCGAGATCCAGAGAGCAGG 1657  
 DB 1441 CCGTGAAGCTGTGATCTACGAGCCCGCAGAGCCTGTGGCGCGAGATCCAGAGAGCAGG 1500  
 QY 1658 GCCACCAAGCAGTGGACCTTACAGAGCTTCAAGAGCCCTTCAAGAACCTTGAAGACCGCA 1717  
 DB 1501 GCCAGGCGCAGTGGACCTTACAGAGCTTCAAGAGCCCTTCAAGAACCTTGAAGACCGCA 1560  
 QY 1718 AGTACGCGAAGATGGCGACCGCCACACCAAGAGAGAGCTGAGAGAGAGAGAGAGAGC 1777  
 DB 1561 AGTACGCGCGATGGCGCGCGCCACCAAGAGAGAGCTGAGAGAGAGAGAGAGAGC 1620  
 QY 1778 AGAAGATCGCGATGGAGAGCAGCTGATCTGGGCGCAAGACCCCGCAAGTTCGCGTCCCA 1837  
 DB 1621 AGAAGTGAAGCAGCAGAGAGCAGCTGATCTGGGCGAAGATCCCAAGTTCAGAGTCCCA 1680  
 QY 1838 TCAGAGAGAGAGCCTGGGAGAGCCTGTGTGAGACCGATCTTGGAGAGCAGCTGAGAGAGAGC 1897



[illegible]

Qy	2438	GGGCTAGCACCGGTGAATTC	2457
Db	2281	GGGCTAGCACCGGTGAATTC	2300
RESULT 6			
BD263704			
LOCUS			
DEFINITION	BD263704	Improved expression of HIV polypeptides and production of virus-like particles.	
ACCESSION	BD263704.1	GI:33073472	
VERSION	JP 2002533124-A/71.		
KEYWORDS		synthetic construct	
SOURCE		synthetic construct	
ORGANISM		artificial sequences.	
REFERENCE	1	(bases 1 to 2306)	
AUTHORS	Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.		
TITLE	Improved expression of HIV polypeptides and production of virus-like particles		
JOURNAL	Patent: JP 2002533124-A	71 08-OCT-2002;	
COMMENT	CHIRON CORP		
	OS Artificial Sequence		
	PN JP 2002533124-A/71		
	PD 08-OCT-2002		
	PF 30-DEC-1999	JP 2000591193	
	PR 31-DEC-1998	US 60/114495, 01-DEC-1999	US 60/168471 PI
	SUSAN BARNETT, JAN ZUR MESEDE, INDRSH SRIVASTAVA, YING LIAN, PI KAREN HARTOG,		
	PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER	PC C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC C12N5/10,	
	PC C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC		
	Description of Artificial Sequence: rs(-).protmod.RTopt.YM FH Key		
	Location/Qualifiers		
FT source	1..2306		
FT	/organism='Artificial Sequence'.		
FEATURES			
source	Location/Qualifiers		
	1..2306		
	/organism='synthetic construct'		
	/mol_type='genomic DNA'		
	/db_xref='taxon:32630'		
ORIGIN			
Query Watch	82.2%	Score 2019.2; DB 6; Length 2306;	
Best Local Similarity	93.2%;	Pred.No. 1.5e-202;	
Matches 2150;	Conservative	0; Mismatches 138; Indels 18; Gaps 3;	
Qy	170	GCGGCAAGAGGGCCACCATGAAGAAGTGCACGAGGCCAGGCCAACTCTTCGCG	229
Db	1	GCGGCGCGAAGGACACCAATGAAAGATTGCCTGAGAGCAGGCTAATTCTTCGCG	60
Qy	230	AGGACTTGCTCCCCAGGCGAAGCCCGCGAGTTCCCAGCAGCAGAACCGCGCCA	289
Db	61	AGGACTTGCTTCTTCGAGGGCAAGCCCGCGAGTTTCAGCAGCGAGCAGCCGCGCA	120
Qy	290	ACAGCCCCACAGCGCGAGCTGCAGTGGCGGGG-----ACAACCCCGCAGCGAGG	343
Db	121	ACAGCCCCACCGCGCGAGCTGCAGTGTGGGGGGGCGAGAACCAACGCTTGAGCGAGG	180
Qy	344	CGGCGCGCAGCGCCAGGGCACCCCTG-----AACTTCCCCAGATCATCCCTGTGGCAGC	397
Db	181	CGGCGCGCAGCCGCGACCGTGCAGTTCACTTCCCTCCAGATCATCCCTGTGGCAGC	240
Qy	398	GCCTCCCTGGTAGCATCAAGTGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGCGG	457
Db	241	GCCTCCCTGGTAGCATCAAGTGGCGGCCAGTCAAAGGAGGCCCTGCTGGACACCGCGG	300
Qy	458	CCGACGACACCGTCTCGAGAGATGAGCTGCCCCGGCAAGTGGAAAGCCCAAGATGATCG	517
Db	301	CCGACGACACCGTCTCGAGAGATGAACTGCTCCGGCAAGTGGAAAGCCCAAGATGATCG	360

[illegible]



TITLE		Expression of HIV polypeptides and production of virus-like particles	
JOURNAL		Patent: US 602705-A 82 05-AUG-2003;	
FEATURES		Location/Qualifiers	
source		1..2306	
ORIGIN		/organism="unknown"	
		/mol_type="genomic DNA"	
Query Match	82.2%;	Score 2019.2;	DB 6; Length 2306;
Best Local Similarity	93.2%;	Pred. No. 1.5e-202;	
Matches 2150;	Conservative	0; Mismatches 138;	Indels 18; Gaps 3;
QY	170	GGCGAAGGAGGCGCCACAGATGAAGAGTGTGACCGAGCGCCAGCGCAACTCTTCCGGG	229
DB	1	GGCGCGCGAAGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGGG	60
QY	230	AGGACCTGGCTTCCCGCCAGCGAAGCCCGCGAGTTCCCGAGCGAGCAAGCCCGCA	289
DB	61	AGGACCTGGCTTCCCGCGAGCGAAGCCCGCGAGTTTACAGCGAGAGACCCGGCGCA	120
QY	290	ACAGCCCAACAGCCCGAGCTCGAGTGGCGGCGGCGG-----ACAAACCCCGAGCGAGG	343
DB	121	ACAGCCCAACAGCCCGAGCTCGAGTGGCGGCGGCGGAGCAACACAGCCTGAGCGAGG	180
QY	344	CCGCGCGAGCGCCAGGCAACCTG-----AACTTCCCGAGATCACCTGTGGCAGC	397
DB	181	CCGCGCGAGCGCCAGGCAACCTGTAGCTTCACTTCCCGAGATCACTGTGGCAGC	240
QY	398	GCCCGCTGTGAGCATCAAGGTGGCGGCGCAGATCAAGAGGCGCTGTGAGACACCGGG	457
DB	241	GCCCGCTGTGATCAATCAGGATCGGCGGCGAGCTCAAGAGGCGCTGTGACACCGGG	300
QY	458	CCGAGGACCGGTGTGGAGGATGAGCTGCGCGGCGAGTGGAGGCGGCAAGATGATCG	517
DB	301	CCGAGGACCGGTGTGGAGGATGAGCTGCGCGGCGAGTGGAGGCGGCAAGATGATCG	360
QY	518	GCAGCATCGCGGCTTCAATCAAGGTGGCGGCGCAGTACGACAGATCCTGATCGAGAT	577
DB	361	GCAGCATCGCGGCTTCAATCAAGGTGGCGGCGCAGTACGACAGATCCTGATCGAGAT	420
QY	578	GCAAGAGGCGATCGGACCGTGTGATCGCGCGGCGAGTGGAGGCGGCAAGATCGGCGCA	637
DB	421	GCAAGAGGCGATCGGACCGTGTGATCGCGCGGCGAGTGGAGGCGGCAAGATCGGCGCA	480
QY	638	ACATGTGACCCAGCTGGGTGACCTGAACTTCCCGATTCGCGCCCATCGAGACCGGTG	697
DB	481	ACCTGTGACCCAGATCGGTGACCTGAACTTCCCGATTCGCGCCCATCGAGACCGGTG	540
QY	698	CCGTGAAGCTGAAGCCCGGATGGAACCGCGGCGAGGTGAAGAGTGGCCCTGACCGAGG	757
DB	541	CCGTGAAGCTGAAGCCCGGATGGAACCGCGGCGAGGTGAAGAGTGGCCCTGACCGAGG	600
QY	758	AGAGATCAAGGCGCTGACCGCATCTGCGAGGAGATGGAGAGGAGGCGCAAGATCACCA	817
DB	601	AGAGATCAAGGCGCTGAGGATCTGACCGGATGGAAGAGGAGGCGCAAGATCAGCA	660
QY	818	AGATCGGCGCGAGAAACCCCTACAAACCCCGGTGTTCCCGCATCAAGAGAGGAGCA	877
DB	661	AGATCGGCGCGAGAAACCCCTACAAACCCCGGTGTTCCCGCATCAAGAGAGGAGCA	720
QY	878	CCAAGTGGCGAAGCTGTGAGTTCGCGAGCTGAACAGCGCACCGAGACTTCTGGG	937
DB	721	CCAAGTGGCGAAGCTGTGAGTTCGCGAGCTGAACAGCGCACCGAGACTTCTGGG	780
QY	938	AGTGTGAGTGGGATCCCCCAACCCCGCGGCTGGAAGAGAGAGAGCGTGAACCGTGC	997
DB	781	AGTGTGAGTGGGATCCCCCAACCCCGCGGCTGGAAGAGAGAGAGCGTGAACCGTGC	840
QY	998	TGGACGTGGGCGAGCTTACTTACGCTGCCCTGGAGAGGACTTCCGCAAGTACACCG	1057
DB	841	TGGACGTGGGCGAGCTTACTTACGCTGCCCTGGAGAGGACTTCCGCAAGTACACCG	900

QY	1058	CCTTACCATCCCGAGATCAACAGAGACCCCGGATCCGCTACAGTACAGTGC	1117
DB	901	CCTTACCATCCCGAGATCAACAGAGACCCCGGATCCGCTACAGTACAGTGC	960
QY	1118	TGCCCCAGGCTGGAAGGGCAGCCCGAGATCTTCCAGAGCAGCAGTACCAAGATCCTGG	1177
DB	961	TGCCCCAGGCTGGAAGGGCAGCCCGGATCTTCCAGAGCAGCAGTACCAAGATCCTGG	1020
QY	1178	AGCCCTTCCGCGCGCAACCCCGAGATCGTGAATACAGGAGCCCGCTGTAGTGGCA	1237
DB	1021	AGCCCTTCCGCAAGCAGAACCCCGAGATCGTGAATACAGGAGCCCGCTGTAGTGGCA	1080
QY	1238	GCAGCTTGGAGATCGGCGAGCAGCCGCGCAAGATCCAGAGCTGCGCAGCAGCCTGCTGC	1297
DB	1081	GCAGCTTGGAGATCGGCGAGCAGCCGCAAGATCCAGAGCTGCGCAGCAGCCTGCTGC	1140
QY	1298	GCTGGGCTTCCACACCCCGCAAGAGCAACAGAGGAGCCCGCTTCTCCCAT--	1355
DB	1141	GCTGGGCTTCCACACCCCGCAAGAGCAACAGAGGAGCCCGCTTCTCTGTGATGG	1200
QY	1356	-----CGAGCTGCACCCCGCAAGTGGAGCGTGCAGCCCATCGAGCTGCCGAGAGGAGA	1411
DB	1201	GCTACGAGCTGCACCCCGCAAGTGGAGCCGTGAGCCCATCATGTGCCCGAGAGGACA	1260
QY	1412	GCTGGAACCTGAAACGACATCCAGAAAGCTGTGGGCAAGCTGAACTGGGCGCAGCATCT	1471
DB	1261	GCTGGAACCTGAAACGACATCCAGAAAGCTGTGGGCAAGCTGAACTGGGCGCAGCATCT	1320
QY	1472	ACCCCGCATCAAGTGTGCGCAGCTGTGCAAGCTGTGCGCGGCGCAAGCCCTGACCG	1531
DB	1321	ACCCCGCATCAAGTGTGCGCAGCTGTGCAAGCTGTGCGCGGCGCAAGCCCTGACCG	1380
QY	1532	ACATGTGCGCCCTGACCGAGGAGCGGAGCTGGAGCTGGCCGAGAACCCGAGATCCTGC	1591
DB	1381	AGGTGATCCCGCTGACCGAGGAGCGGAGCTGGAGCTGGCCGAGAACCCGAGATCCTGA	1440
QY	1592	CGAGCCCGTGCACCGGCTGTACTACGACCCGAGAGGACCTGTGGCGGAGATCCAGA	1651
DB	1441	AGGAGCCCGTGCACCGGCTGTACTACGACCCGAGAGGACCTGTGGCGGAGATCCAGA	1500
QY	1652	AGCAGGCGCACGACAGTGGACCTTACAGATCTACAGAGGAGCCCTTCAAGAACCTGAAGA	1711
DB	1501	AGCAGGCGCACGAGGCTGTGACCTTACAGATCTACAGAGGAGCCCTTCAAGAACCTGAAGA	1560
QY	1712	CCGCAAGTACGCAAGATGGCGACCCGCGCACCAACAGAGCTGAAGCAGCTGACCGAGG	1771
DB	1561	CCGCAAGTACGCGCGCATGCGCGCGCCACCAACAGAGCTGAAGCAGCTGACCGAGG	1620
QY	1772	CCGTGCAAGATCGCCATGGAGAGCATCGTGTGATCTGGGCGAGACCCCAAGTTCGCGC	1831
DB	1621	CCGTGCAAGAGTGAAGCAGCAGAGAGCATCGTGTGATCTGGGCGAGATCCCAAGTTCAGC	1680
QY	1832	TGCCCCATCCAGAGAGACCTGGGAGACCTGTGGAGCCGACTACTGCGAGGCCACCTGGA	1891
DB	1681	TGCCCCATCCAGAGAGACCTGGGAGGCTGTGGATGGAGTACTGGCAGGCCACCTGGA	1740
QY	1892	TCCCGAGTGGAGTCTGTGAACACCCCGCGCTGTGAGCTGTGTGTACCGCTGGAGA	1951
DB	1741	TCCCGAGTGGAGTCTGTGAACACCCCGCGCTGTGAGCTGTGTGTACCGCTGGAGA	1800
QY	1952	AGGAGCCCATCATCGGCGCGAGACCTTCTACGTGAGCGGCGCGCCCAACCGCGAGACCA	2011
DB	1801	AGGAGCCCATCATCGGCGCGAGACCTTCTACGTGAGCGGCGCGCCCAACCGCGAGACCA	1860
QY	2012	AGATCGGCAAGCGGCTACGTGACCGGCGCGGCGCGAGAGATCGTGAAGCTGACCG	2071
DB	1861	AGCTGGGCAAGCGGCTACGTGACCGGCGCGGCGCGAGAGTGTGTGTGAGCATCGCG	1920
QY	2072	AGACCAACCAACAGAGACCGAGCTGACGCGCATCCAGCTGGCGCTGAGAGCAGCGGCA	2131
DB	1921	ACACCAACCAACAGAGACCGAGCTGACGCGCATCCACCTGGCGCTGAGAGCAGCGGCG	1980
QY	2132	GGAGGTGAACATCGTGAACCGAGCAGTACGCGCTGGGCTCATCCAGGCCAGCGCG	2191





Db 4340 ACAGTGGACCGTGCAGCCCATCGTCTGCCGAGAGGACAGCTGACCGTGAACGACA 4399  
Qy 1430 TCAGAGAGCTGGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCGGCATCAAGGTGC 1489  
Db 4400 TCAGAGAGCTGGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCGCGGCATCAAGGTGC 4459  
Qy 1490 GCCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCGCTGACCGACATCGTCCCGCTGACCG 1549  
Db 4460 GCCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCGCTGACCGACATCGTCCCGCTGACCG 4519  
Qy 1550 AGGAGCGGAGCTGAGCTGGCCGAGAACCGCGAGATCTTCCGCGGCCCGTGCACGGCG 1609  
Db 4520 AGGAGCGGAGCTGAGCTGGCCGAGAACCGCGAGATCTTCCGAGAGCGCGTGCACGGCG 4579  
Qy 1610 TGTACTACGACCCAGCAGACCTGTGTGGCCGAGATCCAGAGCAGGCGCCACGACCACT 1669  
Db 4580 TGTACTACGACCCAGCAGACCTGTGTGGCCGAGATCCAGAGCAGGCGCCACGACCACT 4639  
Qy 1670 GGACCTACAGATCTACAGAGGCGCTTCAAGAACCTGTAGACCGGCAAGTACGCCAGA 1729  
Db 4640 GGACCTACAGATCTACAGAGGCGCTTCAAGAACCTGTAGACCGGCAAGTACGCCCGCA 4699  
Qy 1730 TGGCAGCGCCACACCAAGCAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCA 1789  
Db 4700 TGAAGGGCGCCACACCAAGCAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCA 4759  
Qy 1790 TGGAGAGCATCGTGTGAGTGGGCGAAGACCCCAAGTTCGCGCTGCCCATCCAGAGAGA 1849  
Db 4760 CCGAGAGCATCGTGTGAGTGGGCGAAGACCCCAAGTTCGCGCTGCCCATCCAGAGAGA 4819  
Qy 1850 CCTGGGAGACCTGTGAGCAGCAGTACTGGGAGGCGACCTGATCCCGAGTGGAGTTCG 1909  
Db 4820 CCTGGGAGACCTGTGAGCAGCAGTACTGGGAGGCGACCTGATCCCGAGTGGAGTTCG 4879  
Qy 1910 TGAACACCCCGCTGTGAGTGGTGTACAGCTGGAGAGGAGCCCATCATCGCGC 1969  
Db 4880 TGAACACCCCGCTGTGAGTGGTGTACAGCTGGAGAGGAGCCCATCATCGCGC 4939  
Qy 1970 CCGAGACCTTCTAGTGAACGCGCGCCACCGCGAGACCAAGATCGGCAAGCGCGCT 2029  
Db 4940 CCGAGACCTTCTAGTGAACGCGCGCCACCGCGAGACCAAGTGGGCAAGCGCGCT 4999  
Qy 2030 ACGTGACCGACCGGCGCGCGAGAGATCGTGAAGCTGTGACCTGACCGACCAACAGAGA 2089  
Db 5000 ACGTGACCGACCGGCGCGCGAGAGATCGTGAAGCTGTGACCTGACCGACCAACAGAGA 5059  
Qy 2090 CCGAGCTGAGGCACTCAGCTGGCCCTGACGACAGCGCGAGCGAGTGAACATCGTA 2149  
Db 5060 CCGAGCTGAGGCACTCAGCTGGCCCTGACGACAGCGCGCGCTGGAGGTGAACATCGTA 5119  
Qy 2150 CCGAGAGCAGTACGCTTGGGCGATCATCCAGGCGCCAGCCGACAGAGCGAGACCGAGC 2209  
Db 5120 CCGAGAGCAGTACGCTTGGGCGATCATCCAGGCGCCAGCCGACAGAGCGAGCGAGC 5179  
Qy 2210 TGGTGAACAGATCATCAGAGCTGATCAAGAGGAGAGGTGTACTGAGCTGGGTGC 2269  
Db 5180 TGGTGAACAGATCATCAGAGCTGATCAAGAGGAGAGGTGTACTGAGCTGGGTGC 5239  
Qy 2270 CCGCCACAGAGGCGATCGCGCGACAGAGCAGATCGACAACTGGTGAAGAGGCGATCC 2329  
Db 5240 CCGCCACAGAGGCGATCGCGCGACAGAGCAGGTGAGCGGCTGGTGAAGCGCGGATCC 5299  
Qy 2330 GCAAGGTGTGTTCTGAGCGGATCGATGGCGGATCGTGTATCTACAGTA 2381  
Db 5300 GCAAGGTGTGTTCTGAGCGGATCGACAAAGGCGCCAGGAGGAGCAGAGAA 5351

RESULT 9  
BD263706  
LOCUS  
DEFINITION  
ACCESSION

2312 bp DNA linear PAT 17-JUL-2003

Improved expression of HIV polypeptides and production of virus-like particles.

BD263706

BD263706.1 GI:33073474  
JP 2002533124-A/73.  
synthetic construct  
synthetic construct  
artificial sequences.  
1 (bases 1 to 2312)  
Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.  
Improved expression of HIV polypeptides and production of virus-like particles  
Patent: JP 2002533124-A 73 08-OCT-2002;  
CHIRON CORP  
OS Artificial Sequence  
FN JP 2002533124-A/73  
PD 08-OCT-2002  
PF 30-DEC-1998 JP 2000591193  
PR 31-DEC-1998 US 60/114495.01-DEC-1999 US 60/168471 PI  
SUSAN BARNETT, JAN ZUR MEDEDE, INDRESH SRIVASTAVA, YING LIAN, PI  
KARIN HARTOG  
PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER, PC  
C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC  
C12N5/10,  
PC C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC  
Description of Artificial Sequence: FS(-).protmod.RTopt(+) FH Key  
Location/Qualifiers  
FT source 1..2312  
FT /organism='Artificial Sequence'.  
Location/Qualifiers  
1..2312  
/organism='synthetic construct'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32630'  
ORIGIN  
Query Match 81.3%; Score 1998.4; DB 6; Length 2312;  
Best Local Similarity 92.9%; Pred. No. 2.3e-200;  
Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;  
Qy 170 GCGGCAAGGAGGCGCCACAGATGAAGGAGCTGCACGCGCGCAGCCCAACTTCTTCCCG 229  
Db 1 GCGGCGGAGAGACACCAATGAAGATTGCATGAGAGACAGGCTAATTTCTTCCGCG 60  
Qy 230 AGGACTGGCTTCCCGCAGGCGCGCGAGTTCGCCAGCGAGCAAGCCGCGCA 289  
Db 61 AGGACTGGCTTCCCGCAGGCGCGCGAGTTCGCCAGCGAGCAAGCCGCGCA 120  
Qy 290 ACAGCCCGACCGCGCGAGCTGCAGGTGCGCGCGG-----ACAAACCCCGCGAGCGAGG 343  
Db 121 ACAGCCCGACCGCGCGAGCTGCAGGTGCGCGCGGCGAGACAAACAGCTGAGCGAGG 180  
Qy 344 CCGGCGCGAGCGCGCGAGCGACCCCTG-----AAGTCCCGAGATCAAGCTTGGCAGC 397  
Db 181 CCGGCGCGAGCGCGCGAGCGACCCGTGAGCTTCACTCCCGCAGATCAAGCTTGGCAGC 240  
Qy 398 GCGGCTGTGAGCATCAAGTGGCGCGCGAGATCAAGGAGCGCGCTGCTGACACCGCGCG 457  
Db 241 GCGGCTGTGAGCATCAAGTGGCGCGCGAGTTCAGGAGCGCGCTGCTGACACCGCGCG 300  
Qy 458 CCGAGACACCGTGTGAGGAGATGAGCTTCCCGCGCAAGTGGAGCGCGCAAGATGATCG 517  
Db 301 CCGAGACACCGTGTGAGGAGATGAGCTTCCCGCGCAAGTGGAGCGCGCAAGATGATCG 360  
Qy 518 GCGGCTGTGAGCATCAAGTGGCGCGCGAGTTCAGAGTGCAGATCCTGATCCGAGATCTCG 577  
Db 361 GCGGCTGTGAGCATCAAGTGGCGCGCGAGTTCAGAGTGCAGATCCTGATCCGAGATCTCG 420  
Qy 578 GCAAGAGCGCATCGGCA 637  
Db 421 GCGCGAGCGCATCGGCA 480  
Qy 638 ACATGTGACCGAGCTGGGCTGCACCTGAGTTCCTCCATCAGCCCGCGCGCGCGCGCGCG 697  
Db 481 ACCTGCTGACCGAGATCGGCTGCACCTGAGTTCCTCCATCAGCCCGCGCGCGCGCGCG 540

VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BD263706.1 GI:33073474  
JP 2002533124-A/73.  
synthetic construct  
synthetic construct  
artificial sequences.  
1 (bases 1 to 2312)  
Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.  
Improved expression of HIV polypeptides and production of virus-like particles  
Patent: JP 2002533124-A 73 08-OCT-2002;  
CHIRON CORP  
OS Artificial Sequence  
FN JP 2002533124-A/73  
PD 08-OCT-2002  
PF 30-DEC-1998 JP 2000591193  
PR 31-DEC-1998 US 60/114495.01-DEC-1999 US 60/168471 PI  
SUSAN BARNETT, JAN ZUR MEDEDE, INDRESH SRIVASTAVA, YING LIAN, PI  
KARIN HARTOG  
PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER, PC  
C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC  
C12N5/10,  
PC C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC  
Description of Artificial Sequence: FS(-).protmod.RTopt(+) FH Key  
Location/Qualifiers  
FT source 1..2312  
FT /organism='Artificial Sequence'.  
Location/Qualifiers  
1..2312  
/organism='synthetic construct'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32630'

QY 698 CCCTGAAGCTGAAGCCCGGCAATGACGCGCCCAAGGTGAAGCACTGGCCCTTGAACGAGG 757  
Db 541 CCCTGAAGCTGAAGCCCGGATGAGCGCCCAAGGTCAAGCACTGGCCCTTGAACGAGG 600  
QY 758 AGAAGATCAAGCCCTGACCCCATCTGCGAGGAGATGGAAGAGGCGCAAGATCACCA 817  
Db 601 AGAAGATCAAGCCCTGCTGGAGATCTGCAACGAGATGGAAGAGGCGCAAGATCAGCA 660  
QY 818 AGATCGGCCCGAGAACCCCTACAAACACCCCGCTGTTCCGCATCAAGAAGAAGACAGCA 877  
Db 661 AGATCGGCCCGAGAACCCCTACAAACACCCCGCTGTTCCGCATCAAGAAGAAGACAGCA 720  
QY 878 CCAAGTGGCGCAAGCTGTGATCTTCCGCGAGCTGAACAACGCGCACCCAGCACTTCTGGG 937  
Db 721 CCAAGTGGCGCAAGCTGTGATCTTCCGCGAGCTTCCGCGAGCTGAACAACGCGCACCCAGCACTTCTGGG 780  
QY 938 AGGTGAGCTGGGCATCCCGACCCCGCGCGCTGAAGAAGAAGACGCTGACCGTGC 997  
Db 781 AGGTGAGCTGGGCATCCCGACCCCGCGCGCTGAAGAAGAAGACGCTGACCGTGC 840  
QY 998 TGAAGCTGGCGCAAGCTGTGATCTTCCGCGAGCTGAACAACGCGCACCCAGCACTTCTGGG 1057  
Db 841 TGAAGCTGGCGCAAGCTGTGATCTTCCGCGAGCTTCCGCGAGCTGAACAACGCGCACCCAGCACTTCTGGG 900  
QY 1058 CCTTCAACCATCCCGAGCATCAACACGAGACCCCGCGCATCCGCTTACCAAGTACCAACGTCG 1117  
Db 901 CCTTCAACCATCCCGAGCATCAACACGAGACCCCGCGCATCCGCTTACCAAGTACCAACGTCG 960  
QY 1118 TGCCCGAGGCTGGAAGGCGAGCCCGAGCATCTTCCGAGCGAGCATGACCAAGTACCTGG 1177  
Db 961 TGCCCGAGGCTGGAAGGCGAGCCCGCGCATCTTCCGAGCGAGCATGACCAAGTACCTGG 1020  
QY 1178 AGCCCTTCCGCGCGCGCAACCCCGAGATCGTATCTTACCAAGTACCAAGTACCTGG 1231  
Db 1021 AGCCCTTCCGCGCGCGCAACCCCGAGATCGTATCTTACCAAGTACCAAGTACCTGG 1080  
QY 1232 TGGCAGCGACCTGGAGATCGCGCGAGCCCGCAAGATCGAGAGCTGGCGCAAGCACC 1291  
Db 1081 TGGCAGCGACCTGGAGATCGCGCGAGCCCGCAAGATCGAGAGCTGGCGCGCAGCACC 1140  
QY 1292 TGTGGCTGGGGCTTACACCCCGCAAGAGCAGCCAGAGAGCCCGCTTCTCTG 1351  
Db 1141 TGTGGCTGGGGCTTACACCCCGCAAGAGCAGCCAGAGAGCCCGCTTCTCTG 1200  
QY 1352 CCAT-----CGAGCTGCAACCCCGCAAGTGGACCGTGCAGCCCATCGAGCTGCGCGAGA 1405  
Db 1201 GGATGGCTACGAGCTGCAACCCCGCAAGTGGACCGTGCAGCCCATCGAGCTGCGCGAGA 1260  
QY 1406 AGGAGAGCTGACCGTGAACGACATCCAGAGCTGGTGGCGAAGCTGAAGTGGCGCAGCC 1465  
Db 1261 AGGAGAGCTGACCGTGAACGACATCCAGAGCTGGTGGCGAAGCTGAAGTGGCGCAGCC 1320  
QY 1466 AGATCTACCCCGGCATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCC 1525  
Db 1321 AGATCTACCCCGGCATCAAGGTGAAGAGCTGTGCAAGCTGTGCGCGCGCAAGGCC 1380  
QY 1526 TGACCGACATCTGTCCTTGAACGAGCGAGCTGGAGCTGGCGCGAGAACCGCGAGA 1585  
Db 1381 TGACCGAGGTATCTCCCTTGAACGAGCGAGCTGGAGCTGGCGCGAGAACCGCGAGA 1440  
QY 1586 TCTTGGCGCGAGCCCGTGCAGCGGTGTATTAAGACCCCGCAAGAGCACTGGTGGCGCGAGA 1645  
Db 1441 TCTTGAAGAGCGCCGTGCAGCGGTGTATTAAGACCCCGCAAGAGCACTGGTGGCGCGAGA 1500  
QY 1646 TCCAGAGAGCGGCGCAGCGAGTGGACCTTACAGATCTTACAGGAGCCCTTCAAGAAC 1705  
Db 1501 TCCAGAGAGCGGCGCAGCGAGTGGACCTTACAGATCTTACAGGAGCCCTTCAAGAAC 1560  
QY 1706 TGAAGACCGGCAAGTACCGCAAGATGCGCACCGCCCAACCAACGACGAGTGAAGAGCTGA 1765  
Db 1561 TGAAGACCGGCAAGTACCGCCCGCATGCGCGCGCCCAACCAACGACGAGTGAAGAGCTGA 1620

QY 1766 CCGGCGCGTGCAGAGATCGCCATGGAGAGCATCGTGTATCTGGGGCAAGACCCCAAGT 1825  
Db 1621 CCGAGGCGCTGCGAAGGTGAGCACCAGAGCATCTGTATCTGGGGCAAGATCCCAAGT 1680  
QY 1826 TCCGCTCTGCCATCCAGAAAGAGAGCCTGGGAGA CTTGTGTGACCGACTACTTGGCAGGCA 1885  
Db 1681 TCAAGCTGCCATCCAGAAAGAGAGCCTGGGAGGCTTGTGTGATGGAGTACTTGGCAGGCA 1740  
QY 1886 CTTGTGATCCCGAGTGGGAGTTCGTGAACACCCCGCTTGTGAAGCTGTGTGATCCAGC 1945  
Db 1741 CTTGTGATCCCGAGTGGGAGTTCGTGAACACCCCGCTTGTGAAGCTGTGTGATCCAGC 1800  
QY 1946 TGAAGAAGAGCGCCATCATTCGCGCGAGAGCCTTCTACGTGAGCGCGCGCCCAACCGG 2005  
Db 1801 TGAAGAAGAGCGCCATCATTCGCGCGAGAGCCTTCTACGTGAGCGCGCGCCCAACCGG 1860  
QY 2006 AGACCAAGATCGGCAAGCGCGCTAGTACCGAGCGCGCGCGCGAGAGATCTGTGAGC 2065  
Db 1861 AGACCAAGATCGGCAAGCGCGCTAGTACCGAGCGCGCGCGCGAGAGATCTGTGAGC 1920  
QY 2066 TGACCGAGACCCCAACCAAGAGACCGAGCTGTCAGGCCCATTCAGTGGCCCTTGCAGGACA 2125  
Db 1921 TCGCCGACACCCCAACCAAGAGACCGAGCTGTCAGGCCCATTCAGTGGCCCTTGCAGGACA 1980  
QY 2126 GCGGCGAGAGGTGAACATCTGTGACGACGAGCTAGCCCTGGGCGCATCTCCAGGCC 2185  
Db 1981 GCGGCGTGAAGTGAACATCTGTGACGACGAGCTAGCCCTGGGCGCATCTCCAGGCC 2040  
QY 2186 AGCCCGACAAAGAGAGAGAGAGCTGTCAGGCCCATTCAGTGGCCCTTGCAGGACA 2245  
Db 2041 AGCCCGACAAAGAGAGAGAGCTGTCAGGCCCATTCAGTGGCCCTTGCAGGACA 2100  
QY 2246 AGAAGTGTACTGTGAGCTGGTGGCGCGCGCGAGCGCATCGCGCGAGCAGCAGTGG 2305  
Db 2101 AGAAGTGTACTGTGAGCTGGTGGCGCGCGCGAGCGCATCGCGCGAGCAGCAGTGG 2160  
QY 2306 ACAAGCTGTGAGAGAGGCGATCCGAAAGTGTCTTCTGAGCGCGCATCGATGGCGGCA 2365  
Db 2161 ACAAGCTGTGAGAGGCGCGCATCCGAAAGTGTCTTCTGAGCGCGCATCGATGGCGGCA 2220  
QY 2366 TCGTGTACTACAGTACATGAGCAGCTGTGACGTGGGAGCGCGCGCGCTAGGATCGATT 2425  
Db 2221 TCGTGTACTACAGTACATGAGCAGCTGTGACGTGGGAGCGCGCGCGCTAGGATCGATT 2280  
QY 2426 AAAAGCTTCCGGGGCTAGCACCGGTGAATTC 2457  
Db 2281 AAAAGCTTCCGGGGCTAGCACCGGTGAATTC 2312

RESULT 10  
AR373389 AR373389 2312 bp DNA linear PAT 18-DEC-2003  
LOCUS Sequence 84 from patent US 6602705.  
DEFINITION AR373389  
ACCESSION AR373389  
VERSION AR373389.1 GI:40075492  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2312)  
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.  
TITLE Expression of HIV polypeptides and production of virus-like particles  
JOURNAL Patent: US 6602705-A 84 05-AUG-2003;  
FEATURES Location/Qualifiers  
1..2312  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 81.3%; Score 1998.4; DB 6; Length 2312;  
Best Local Similarity 92.9%; Pred. No. 2.3e-200;  
Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;

QY 170 GCGCAAGGAGGCGCACAGATGAAGACTGACCGAGCGCCAGGCGCAACTTCTTCGGG 229  
Db 1 CGGCGCGAAGGACACAAATGAATGCACTGAGAGACAGGCTAATTTCTTCGGCG 60  
QY 230 AGGACTTGCGCTTCCCGCAGGCGAAGCGCGAGTTCCCGAGCGAGAGACCGCGCCA 289  
Db 61 AGGACTTGCGCTTCCCGCAGGCGAAGCGCGAGTTCCCGAGCGAGAGACCGCGCCA 120  
QY 290 ACAGCCCAACAGCGCGAGCTGCAGGTGCGCGCG- - - - -ACRACCCCGCGAGG 343  
Db 121 ACAGCCCAACCGCGCGAGCTGCAGGTGTTGGGCGCGAGAACAAACGCTGAGCGAG 180  
QY 344 CGGCGCGAGCGCGACCGCTG- - - - -AATTCGCCAGATCAACCTGTGCGACG 397  
Db 181 CGGCGCGAGCGCGACCGCTGAGCTTCACTTCCCGCAGATCAACCTGTGCGACG 240  
QY 398 GCCCGCTGCTGAGCATCAAGTGGCGCGCGAGATCAAGAGGCGCTGCTGGACACCGCG 457  
Db 241 GCCCGCTGCTGACCATCAGATCGCGCGCGAGCTCAAGAGGCGCTGCTGGACACCGCG 300  
QY 458 CCGACGACACCGTCTGAGAGAGATGAGCTGCGCGCGAAGTGGAAAGCCCAAGATGTCG 517  
Db 301 CCGACGACACCGTCTGAGAGAGATGAACTGCGCGCGAAGTGGAAAGCCCAAGATGTCG 360  
QY 518 GCGGCATCGCGCGCTTCATCAAGTGGCGCGCGAGTACGACAGATCTCTGATCGAGATCTCG 577  
Db 361 GCGGCATCGCGCGCTTCATCAAGTGGCGCGCGAGTACGACAGATCTCTGATCGAGATCTCG 420  
QY 578 GCAAGAGGCGCATCGGACCGTGTGATCGGCGCGCACCGCGTGAACATCATCGGCGCGCA 637  
Db 421 GCCAAGAGGCGCATCGGACCGTGTGATCGGCGCGCACCGCGTGAACATCATCGGCGCGCA 480  
QY 638 ACATGCTGACCGAGCTGGGCTGACACCTGAACTTCCCGCATCAGCGCGCATCGAGACCGTGC 697  
Db 481 ACCTGCTGACCGAGCTGGGCTGACACCTGAACTTCCCGCATCAGCGCGCATCGAGACCGTGC 540  
QY 698 CCGTGAAGCTGAGACCGCGCATGAGCGCGCGCGAGTGAAGTGAAGTGGCGCGCTGACCGAGG 757  
Db 541 CCGTGAAGCTGAGACCGCGCATGAGCGCGCGCGAGTGAAGTGAAGTGGCGCGCTGACCGAGG 600  
QY 758 AGAAGATCAAGGCGCTGACCGCATCTGCGAGGAGATGGAGAGGAGGCGAAGATCAACA 817  
Db 601 AGAAGATCAAGGCGCTGAGAGATCTGCACCGAGATGGAGAGGAGGCGAAGATCAACA 660  
QY 818 AGATCGCGCGCGAGACCGCTACACACCGCGCTGTCGCATCAGAGAGAGACAGCA 877  
Db 661 AGATCGCGCGCGAGACCGCTACACACCGCGCTGTCGCATCAGAGAGAGACAGCA 720  
QY 878 CCAAGTGGCGCAAGCTGCTGGAATTCGCGAGCTGAAACAGCGCGACCGCAGGACTTCTGGG 937  
Db 721 CCAAGTGGCGCAAGCTGCTGGAATTCGCGAGCTGAAACAGCGCGACCGCAGGACTTCTGGG 780  
QY 938 AGGTGAGCTGGGCTATCCCGACCGCGCGCTGAGAGAGAGAGAGCGTGAACGCTGC 997  
Db 781 AGGTGAGCTGGGCTATCCCGACCGCGCGCTGAGAGAGAGAGAGCGTGAACGCTGC 840  
QY 998 TGAAGTGGCGCAAGCTGCTGGAATTCGCGAGCTGAAACAGCGCGACCGCAGGACTTCTGGG 1057  
Db 841 TGAAGTGGCGCAAGCTGCTGGAATTCGCGAGCTGAAACAGCGCGACCGCAGGACTTCTGGG 900  
QY 1058 CTTTACCATCCCGAGCATCAACAGAGACCGCGCGCTGAGAGAGAGAGAGCGTGAACGCTGC 1117  
Db 901 CTTTACCATCCCGAGCATCAACAGAGACCGCGCGCTGAGAGAGAGAGAGCGTGAACGCTGC 960  
QY 1118 TGCCCGAGGCTGAGAGGCGAGCGCGCATCTTCAGAGCAGCATGACCAAGATCTCTGG 1177  
Db 961 TGCCCGAGGCTGAGAGGCGAGCGCGCATCTTCAGAGCAGCATGACCAAGATCTCTGG 1020  
QY 1178 AGCCCTTCCCGCGCGCAACCGCGAGATCGTGTATCTACCA- - - - -GGCGCGCGCTGACG 1231  
Db 1021 AGCCCTTCCCGAGCAGACCGCGCATCTGTGTATCTACCGATGATGAGACCGCTGTACG 1080

QY 1232 TGGCAGACGACCTGGAGATCGGCAGCACCGCGCAAGATCTGAGAGCTGCGCAAGCAC 1291  
Db 1081 TGGCAGACGACCTGGAGATCGGCAGCACCGCGCAAGATCTGAGAGCTGCGCAAGCAC 1140  
QY 1292 TGTGCGCTGGGCTTCCACACCGCGCAAGAGACCAAGAGAGAGAGAGAGAGAGAGAGAG 1351  
Db 1141 TGTGCGCTGGGCTTCCACACCGCGCAAGAGAGCAACAGAGAGAGAGAGAGAGAGAGAG 1200  
QY 1352 CCAT- - - - -CGAGTGCACCGCGCAAGTGCAGCTGCGAGCGCATCGAGCTGCGCGAGA 1405  
Db 1201 GATGGGCTACGAGTGCACCGCGCAAGTGCAGCTGCGAGCGCATCATGCTGCGCGAGA 1260  
QY 1406 AGGAGAGCTGGACCGTGAACCGACATCCAGAGCTGCTGGGCAAGCTGAATGGGCGAGAG 1465  
Db 1261 AGGAGAGCTGGACCGTGAACCGACATCCAGAGCTGCTGGGCAAGCTGAATGGGCGAGAG 1320  
QY 1466 AGATCTTACCGCGCATCAAGTGCAGCTGCTGAGCTGCTGCGCGCGCGCGCGCGCGCG 1525  
Db 1321 AGATCTTACCGCGCATCAAGTGCAGCTGCTGAGCTGCTGCGCGCGCGCGCGCGCGCG 1380  
QY 1526 TGAACGACATCTGCTGCGCTGACCGAGAGCGAGCTGAGAGCTGCGCGAGAACCGCGAGA 1585  
Db 1381 TGAACGAGTGTATCCCGCTGACCGAGAGCGAGCTGAGAGCTGCGCGAGAACCGCGAGA 1440  
QY 1586 TCTGCGCGAGCGCTGACCGCGCTGACTACGACCGCGAGAGCTGCTGCGCGCGCGCGCG 1645  
Db 1441 TCTGAGAGAGCGCTGACCGCGCTGACTACGAGCTGACTACGAGAGAGAGCTGCTGCGCGCG 1500  
QY 1646 TCCAGAGAGAGCGCGCACGAGTGGAGCTTACAGATCTTACAGAGAGCGCTTCCAGAAC 1705  
Db 1501 TCCAGAGAGAGCGCGCACGAGTGGAGCTTACAGATCTTACAGAGAGCGCTTCCAGAAC 1560  
QY 1706 TGAAGACCGCGCAAGTACCGCAAGATGCGCACCGCGCGCACCAAGAGAGAGAGAGAG 1765  
Db 1561 TGAAGACCGCGCAAGTACCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1620  
QY 1766 CCGAGCGCTGAGAGAGTCCGCTGAGAGAGCTGAGTCTGGGCGAGAGAGAGAGAGAGAG 1825  
Db 1621 CCGAGCGCTGAGAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
QY 1826 TCGCGCTGCGCATCCAG 1885  
Db 1681 TCAAGCTGCGCATCCAG 1740  
QY 1886 CTTGATCCCGAGTGGAGTGTGAGAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1945  
Db 1741 CTTGATCCCGAGTGGAGTGTGAGAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1800  
QY 1946 TGAAG 2005  
Db 1801 TGAAG 1860  
QY 2006 AGACCAAGATCGGCAAGAGCGCGCTACGTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAG 2065  
Db 1861 AGACCAAGTGGGCAAGAGCGCGCTACGTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920  
QY 2066 TGAACGAG 2125  
Db 1921 TCGCGAGACACCAACAG 1980  
QY 2126 GGGGAG 2185  
Db 1981 GGGGCTGAGAGTGAACATCGTGAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040  
QY 2186 AGCCCGAG 2245  
Db 2041 AGCCCGAG 2100  
QY 2246 AGAGAGTGTACCTGAGCTGGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2305  
Db 2101 AGAGAGTGTACCTGCGCTGGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2160  
QY 2306 ACAAGCTGTGAG 2365





QY 962 CCGCGGCGCTGAAGAAGAGAGCGTGTGACCGTGTGACGCTGGGCGAGCGCCTACTTCA 1021  
Db 3923 CCGCGGCGCTGAAGCAGAGAGAGCGTGTGACCGTGTGACGCTGGGCGAGCGCCTACTTCA 3982  
QY 1022 GCGTGCCTCCGAGCAGAGATTCGCGAAGTACACCGCCTTACCATCCCGAGCATCAACA 1081  
Db 3983 GCGTGCCTCCGAGCAGAGATTCGCGAAGTACACCGCCTTACCATCCCGAGCATCAACA 4042  
QY 1082 ACAGAGACCCCGGATCCGCTACAGTACAGTGTGCTGCCCGAGGCTGGAAGGCGAGCC 1141  
Db 4043 ACAGAGACCCCGGATCCGCTACAGTACAGTGTGCTGCCCGAGGCTGGAAGGCGAGCC 4102  
QY 1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCCGCAACCCCG 1201  
Db 4103 CCGCATCTTCCAGTGCAGCATGACCAAGATCTTGGAGCCCTTCCGCAAGCAGAACCCCG 4162  
QY 1202 AGATCGTGTATACCA-----GCCGCCCTGTAGTGGGCGAGCGACTGTGAGATGGCC 1255  
Db 4163 ACATCGTGTATACCA-----GCCGCCCTGTAGTGGGCGAGCGACTGTGAGATGGCC 4222  
QY 1256 AGCACCGCGCCAGATCGAGAGTGCAGACCTGTGCGTGGGCTTCAACACCC 1315  
Db 4223 AGCACCGCACCAAGATCGAGAGTGCAGACCTGTGCGTGGGCTTCAACACCC 4282  
QY 1316 CCGACAAGACACCAAGAGAGCGCCCTTCTGCCCCAT-----CGAGTGCACCCCG 1369  
Db 4283 CCGACAAGACACCAAGAGAGCGCCCTTCTGCCCCAT-----CGAGTGCACCCCG 4342  
QY 1370 ACAAGTGAACCGTGCAGCCCATCGAGTGCAGAGAGAGAGTGCAGCCGTGAACGACA 1429  
Db 4343 ACAGTGAACCGTGCAGCCCATCGAGTGCAGAGAGAGAGTGCAGCCGTGAACGACA 4402  
QY 1430 TCCAGAGCTGGTGGGCAAGCTGAATGGGCGAGCAGATCTACCCCGGCATCAAGTGC 1489  
Db 4403 TCCAGAGCTGGTGGGCAAGCTGAATGGGCGAGCAGATCTACCCCGGCATCAAGTGC 4462  
QY 1490 GCCAGCTGTGAAGCTGTGCGCGCGCAAGGCGCTGACCGACATCTGTGCCCTGACCG 1549  
Db 4463 GCCAGCTGTGAAGCTGTGCGCGCGCAAGGCGCTGACCGAGTGTGCTCCCTGACCG 4522  
QY 1550 AGGAGCGCAGTGGAGTGCAGAGAACCGCGAGATCTCTGCGAGCCGTGACGCGG 1609  
Db 4523 AGGAGCGCAGTGGAGTGCAGAGAACCGCGAGATCTCTGAGAGCGCCGTGACGCGG 4582  
QY 1610 TGTACTAGACCCCGAGAGAGCTGTGCGCGAGATCCAGAGCAGGCGCCACGACGAGT 1669  
Db 4583 TGTACTAGACCCCGAGAGAGCTGTGCGCGAGATCCAGAGCAGGCGCCAGGCGCAGT 4642  
QY 1670 GGACCTACAGATCTACAGAGAGCGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAGA 1729  
Db 4643 GGACCTACAGATCTACAGAGAGCGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCGCA 4702  
QY 1730 TCGCACCGCCACACCAACGAGCTGAAGCAGTGTGACCGAGCGCGTGCAGAGATCGGCCA 1789  
Db 4703 TGAAGGCGCGCCACACCAACGAGCTGAAGCAGTGTGACCGAGCGCGTGCAGAGATCGGCCA 4762  
QY 1790 TGGAGAGCATCTGTGATCTGGGCGAAGACCCCAAGTTCGCGCTGCCCATCCAGAGAGAGA 1849  
Db 4763 CCGAGAGCATCTGTGATCTGGGCGAAGACCCCAAGTTCAGCTGCCCATCCAGAGAGAGA 4822  
QY 1850 CTGGGAGACTGTGTGAGCAGACTACTGAGAGGCCACTGTGATCCCGAGTGGAGTTCG 1909  
Db 4823 CTGGGAGGCTGTGTGAGCAGACTACTGAGAGGCCACTGTGATCCCGAGTGGAGTTCG 4882  
QY 1910 TGAACACCCCGCCCTGTGTGAGTGTGTATACAGTGTGAGAGAGAGGCCCATCATCGGCG 1969  
Db 4883 TGAACACCCCGCCCTGTGTGAGTGTGTATACAGTGTGAGAGAGAGGCCCATCATCGGCG 4942  
QY 1970 CCGAGACCTTCTACGTGAGCGCGCGCAACCGCGAGACCAAGATTCGCGAGGCGCGCT 2029  
Db 4943 CCGAGACCTTCTACGTGAGCGCGCGCAACCGCGAGACCAAGATTCGCGAGGCGCGCT 5002

QY 2030 ACGTACCGACCGGGCGCGGAGAGATCGTGTGACCTGTGACGAGACCAACCAAGAGA 2089  
Db 5003 ACGTACCGACCGGGCGCGGAGAGTGTGTGCCCTGTGACGAGACCAACCAAGAGA 5062  
QY 2090 CCGAGCTCAGCGCATCCAGCTGCGCTTGCAGGAGCAGCGGAGAGTGAACATCGTGA 2149  
Db 5063 CCGAGCTCAGCGCATCCAGCTGCGCTTGCAGGAGCAGCGGCTGAGGTGAACATCGTGA 5122  
QY 2150 CCGACAGCCAGTACGCGCTTGGGCTATCATCCAGGCCAGCCGCAAAAGCGAGAGCGAGC 2209  
Db 5123 CCGACAGCCAGTACGCGCTTGGGCTATCATCCAGGCCAGCCGCAAAAGCGAGAGCGAGC 5182  
QY 2210 TGGTGAACAGATCATCCAGCAGCTGTATCAAGAGGAGAGTGTGTACTGTAGCTGGGTGC 2269  
Db 5183 TGGTGAACAGATCATCCAGCAGCTGTATCAAGAGGAGAGTGTGTACTGTAGCTGGGTGC 5242  
QY 2270 CCGCCCAACAGGCGCATCGCGCGCAACAGCAGATCGCAACAGTGTGTAGCAAGGCGATCC 2329  
Db 5243 CCGCCCAACAGGCGCATCGCGCGCAACAGCAGAGTGTGTAGCGCTGTGTAGCGCGCGATCC 5302  
QY 2330 GCAAGTGTGTCTTCTGTGACGCGCATCGATGGCGCATCGTGTATCTACCACTA 2381  
Db 5303 GCAAGTGTGTCTTCTGTGACGCGCATCGATGGCGCATCGTGTATCTACCACTA 5354

RESULT 13  
AX427926 9194 bp DNA linear PAT 20-JUN-2002  
LOCUS  
DEFINITION Sequence 164 from Patent WO0232943.  
ACCESSION AX427926  
VERSION AX427926.1 GI:21538013  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM synthetic construct  
1  
REFERENCE Huang, Y. and Nabel, G.J.  
AUTHORS Modifications of hiv env, gag, and pol enhance immunogenicity for  
TITLE Genetic immunization  
JOURNAL Patent: WO 0232943-A 164 25-APR-2002;  
GOVERNMENT OF THE UNITED STATES (US)

FEATURES  
source  
1..9194  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="plasmid pVR1012x/s containing HIV genes"

ORIGIN

Query Match 80.4%; Score 1975; DB 6; Length 9194;  
Best Local Similarity 90.9%; Pred. No. 4.4e-198;  
Matches 2177; Conservative 0; Mismatches 190; Indels 28; Gaps 6;  
QY 14 TGGCCGAGGCCATGAGCCAGG--CCACCAGCGCCAACTCTGTATGACGCGCAGCAACT 70  
Db 2967 TGGCCGAGGCCATGAGCCAGGTTGACCAAGCGCCACCATCATGTGACGCGCGCAACT 3026  
QY 71 TCAAGGCGCCCAAGCCGATCATCAAGTGTTCATCTGCGGCAAGGAGGCGCATCGCCC 130  
Db 3027 TCCGCAACACAGCGCAAGATCGTGAAGTGTTCATCTGCGGCAAGGAGGCGCCAC 3086  
QY 131 GCAACTGCGCGCCCGCCCGCAGAGAGGCTGTGGAAGTGGCGAAGGAGGCGCCACCA 190  
Db 3087 GCAACTGCGCGCGCCCGCCCGCAGAGAGGCTGTGGAAGTGGCGAAGGAGGCGCCACCA 3146  
QY 191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTTCTCCGCGAGGACCTGTGGCTTCCCGCAGG 250  
Db 3147 TGAAGGACTGCACCGAGCGACAGGCTAA-TTTTATAGGGAAGATCTGGCTTCCCAAG 3205  
QY 251 GCAAGGCGCGGAGTTCGCCAGCGAGAGAGAACCGCGCCACAGCCCCACAGCGCGGAGC 310  
Db 3206 GCAAGGCGCGGAGTTCCTTCTCAGAGCAGACAGGCGCCACAGCCCCACAGGAGAGAGC 3265





VERSION	AX427925.1	GI:21538012	
KEYWORDS	synthetic construct		
SOURCE	synthetic construct		
ORGANISM	artificial sequences.		
REFERENCE	1		
AUTHORS	Huang, Y. and Nabel, G. J.		
TITLE	Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization		
JOURNAL	Patent: WO 0232943-A 163 25-APR-2002;		
GOVERNMENT OF THE UNITED STATES (US)	Location/Qualifiers		
FEATURES	1..9194		
source	/organism="synthetic construct"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:32630"		
	/note="plasmid pVR1012x/s containing HIV genes"		
ORIGIN			
Query Match	80.3%;	Score 1973.4;	DB 6; Length 9194;
Best Local Similarity	90.9%;	Pred. No. 6.5e-198;	
Matches 2176;	Conservative	0; Mismatches 191;	Indels 28; Gaps 6;
QY	14	TGCGCGAGCCATGAGCCAGG---CCACGAGCCCAACATCTGTATGAGCGCGCAACT	70
DB	2967	TGCGCGAGCCATGAGCCAGGTCACCAAGCGGCCACCATATGATGAGCGCGCAACT	3026
QY	71	TCAAGGGCCCCAAGCGGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGCCACATCGCCC	130
DB	3027	TCCGCAACAGCGCAAGATCGTGAAGTCTTCAACTGCGGCAAGGAGGCCACACCGGCC	3086
QY	131	GCAATCGCGCGCCCGCCCGCAGAGGGTGTGGAAGTGTGGGCAAGGAGGCCACACAGA	190
DB	3087	GCAATCGCGCGCCCGCCCGCAGAGGGTGTGGAAGTGTGGGCAAGGAGGCCACACAGA	3146
QY	191	TGAAGGACTGCAACCGAGCGCCAGGCAACTTCTTCCGGAGGACCTGGCCCTTCCGCCAGG	250
DB	3147	TGAAGGACTGCAACCGAGCGCACAGGCTAA-TTTTTAGGGAAGATCTGGCCCTCCACAA	3205
QY	251	GCAAGCGCGCGAGTTCACGAGAGCAGAACCGCGCCGACAGCCCGCCACGAGCCCGGAGC	310
DB	3206	GGAAGCGCCAGGGAATTTTCTTCCAGAGCAGCAGAGCCCAACAGCGCCCAACGAGAGAGC	3265
QY	311	TGCAGTGTGCGCG-----CGACAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCA	364
DB	3266	TTCAGGTTTGGGAGAGAGACAACATCCCTCTCAGAGCAGAGCCGATAGACAGGAA	3325
QY	365	-----CCCTGAATTCCTCCAGATCACCTGTGCGAGCGCCCTGTGTGAGCATCAAGG	418
DB	3326	CTGTATCTTTAGTTCCTCTCAGATCACTCTTTGGCAGCGACCTCTGTCAATAAAGA	3385
QY	419	TGGCGCGCAGATCAAGAGGGCTGCTGGACACCGCGCGCGCAGACCGCTGTGAGG	478
DB	3386	TAGGGGGCCAGCTGAAGAGGGGCTTTCTAGACACCGCGCGCGCAGACCGGTGTGGAGG	3445
QY	479	AGATGAGCTGTGCGCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTCATCA	538
DB	3446	AGATGAACCTGCGCGCGCTGGAAGCCCAAGATGATCGCGGCATCGCGGCTTCATCA	3505
QY	539	AGGTGCGCAGTACGACAGATCTCTGATCGAGATCTGGGCAAGAGGCGCATCGGCACG	598
DB	3506	AGGTGCGCAGTACGACAGATCTCTGATCGAGATCTGGGCAAGAGGCGCATCGGCACG	3565
QY	599	TGCTGATCGGCGCCACCCCGGTGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGGCT	658
DB	3566	TGCTGTTGGGCGCCACCCCGGTGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGCT	3625
QY	659	GCACTGTAACTTCCCATCAGCCCATCGAGACCGTGTGCGGTGAGTGAAGCCCGGCA	718
DB	3626	GCACCTGTAACTTCCCATCAGCCCATCGAGACCGTGTGCGGTGAGTGAAGCCCGGCA	3685
QY	719	TGACCGGCCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGCCCTTGACCG	778

3686	TGACCGGCCCAAGGTGAAGCAGTGGCCCTTGACCGAGAGAGATCAAGGCCCTGTGGTG	3745
779	CCATCTGCGAGGAGATGGAGAAGGAGGCAAGATCACCAAGATCGGCCCGCCGAGAACCCCT	838
3746	AGATCTGACCGAGATGGAGAAGGAGGCAAGATCAGCAAGATCGGCCCGAGAACCCCT	3805
839	ACAACACCCCGCTGTTCCTCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	898
3806	ACAACACCCCGCTGTTCCTCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	3865
899	ACTTCCGCGAGCTGAACAGAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCC	958
3866	ACTTCCGCGAGCTGAACAGAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCC	3925
959	ACCCCGCGCGCTGAAG	1018
3926	ACCCCGCGCGCTGAAG	3985
1019	TGAGCGTGGCCCTGGAG	1078
3986	TGAGCGTGGCCCTGGAG	4045
1079	ACAACAGAGACCCCGCGCATCCGCTACAGTACAACGCTGTGCCCCAGAGGCTGGAAGGCA	1138
4046	ACAACAGAGACCCCGCGCATCCGCTACAGTACAACGCTGTGCCCCAGAGGCTGGAAGGCA	4105
1139	GCCCCAGATCTTCCAGAGAGAGATGACCAAGATCTGAGAGCTTCCGCGCGCGCAAC	1198
4106	GCCCCAGATCTTCCAGAGAGATGACCAAGATCTGAGAGCTTCCGCGCGCGCAAC	4165
1199	CCGAGATCGTGTATACCA-----GGCCCCCTGTACGTGGGCGAGCAGCTTGGAGATCG	1252
4166	CCGAGATCGTGTATACCA-----GGCCCCCTGTACGTGGGCGAGCAGCTTGGAGATCG	4225
1253	GCCAGACCGCGCCAGATGAG	1312
4226	GCCAGACCGCGCCAGATGAG	4285
1313	CCCCGAG	1366
4286	CCCCGAG	4345
1367	CCGACAGTGGACCGTGCAGCCCATCGAGTGTGCCCGAGAGAGAGAGAGAGAGAGAGAGAG	1426
4346	CCGACAGTGGACCGTGCAGCCCATCGAGTGTGCCCGAGAGAGAGAGAGAGAGAGAGAGAG	4405
1427	ACATCCAGAGCTGTGGGCGAGCTGAATGGGCGAGCAGATCTACCCCGGATCAAGG	1486
4406	ACATCCAGAGCTGTGGGCGAGCTGAATGGGCGAGCAGATCTACCCCGGATCAAGG	4465
1487	TGCGCCAGCTGTCAAGCTGCTGCGCGCGCCAAAGGCCCTTGACCGACATCGTSCCCTGA	1546
4466	TGCGCCAGCTGTCAAGCTGCTGCGCGCGCCAAAGGCCCTTGACCGAGGTGTGCCCTGA	4525
1547	CCGAGAGGCGAGCTGGAGCTGGCGAGAACCGCGAGATCTTGGCGGAGCCCGTGCAG	1606
4526	CCGAGAGGCGAGCTGGAGCTGGCGAGAACCGCGAGATCTTGGCGGAGCCCGTGCAG	4585
1607	GGGTGTACTACGACCCCGAGCAAGGACTGTGTGGCGCGAGATCCCAAGAGAGAGAGAGAGAG	1666
4586	GGGTGTACTACGACCCCGAGCAAGGACTGTGTGGCGCGAGATCCCAAGAGAGAGAGAGAGAG	4645
1667	AGTGGACCTACAGATCTTACAGAGCCCTTCAAGAACTGTAAGAGAGAGAGAGAGAGAGAGAG	1726
4646	AGTGGACCTACAGATCTTACAGAGCCCTTCAAGAACTGTAAGAGAGAGAGAGAGAGAGAGAG	4705
1727	AGATGCCACCGGCCCAACCAACGAGTGAAGCAGCTGACCGAGGCGGTGCAAGAGATCG	1786
4706	GCATGAAGGCGGCCCAACCAACGAGTGAAGCAGCTGACCGAGGCGGTGCAAGAGATCG	4765
1787	CCATGGAGAGATCGTGTATCTGGGGCAAGACCCCGCAAGTTCCGCTTCCCATCCAGAGG	1846
4766	CCACCGAGAGATCGTGTATCTGGGGCAAGACCCCGCAAGTTCCGCTTCCCATCCAGAGG	4825



QY 1847 AGACCTGGGAGACCTGGTGGACCGACTACTGGGAGGCGCACTGGATCCCGAGTGGAGT 1906  
Db 4826 AGACCTGGGAGGCTGGTGGACCGAGTACTGGCAGCCACCTGGATCCCGAGTGGAGT 4885  
QY 1907 TCSTGAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCG 1966  
Db 4886 TCSTGAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCG 4945  
QY 1967 GCGCCGAGACCTTCTAGCTGGAGCGGCGCGCCAAACCGGAGACCAAGATCGGCAAGCGCG 2026  
Db 4946 GCGCCGAGACCTTCTAGCTGGAGCGGCGCGCCAAACCGGAGACCAAGATCGGCAAGCGCG 5005  
QY 2027 GCTACGTGACCGACCGGCGCGCCAGAGATCGTGAGCCCTGACCGAGACCAACACGAGA 2086  
Db 5006 GCTACGTGACCGACCGGCGCGCCAGAGATCGTGAGCCCTGACCGAGACCAACACGAGA 5065  
QY 2087 AGACCGAGCTGACGGCGATCCAGCTGGCCCTGACGAGACAGCGGCGAGGTGAACATCG 2146  
Db 5066 AGACCGAGCTGACGGCGATCCAGCTGGCCCTGACGAGACAGCGGCGAGGTGAACATCG 5125  
QY 2147 TGACCGACAGCCAGTACGCGCCCTGGGCATCATCCAGGCCCGAGCCGAGAGAGCG 2206  
Db 5126 TGACCGACAGCCAGTACGCGCCCTGGGCATCATCCAGGCCCGAGCCGAGAGAGCG 5185  
QY 2207 AGCTGTGACAGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTACCTGAGCTGGG 2266  
Db 5186 AGCTGTGACAGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTACCTGAGCTGGG 5245  
QY 2267 TGCCCGCCCAAGGGCATCGCGGCAACGAGCAGATCGACAAGTGTGTGAGCAAGGCGCA 2326  
Db 5246 TGCCCGCCCAAGGGCATCGCGGCAACGAGCAGTGTGAGCGCCCTGGTGTGAGCGCGCA 5305  
QY 2327 TCAGCAAGTGTGTCTGAGCGGCGATCGATGGCGGCATCGTGATCTACAGATA 2381  
Db 5306 TCAGCAAGTGTGTCTGAGCGGCGATCGATGGCGGCATCGACAAGGCCGAGGAGGACCGAGAA 5360

RESULT 15  
AX427927  
LOCUS 12411 bp DNA linear PAT 20-JUN-2002  
DEFINITION Sequence 165 from Patent W00232943.  
ACCESSION AX427927  
VERSION AX427927.1 GI:21538014  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1  
AUTHORS Huang, Y. and Nabel, G. J.  
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization  
JOURNAL Patent: WO 0232943-A 165 25-APR-2002;  
GOVERNMENT OF THE UNITED STATES (US)  
FEATURES  
Location/Qualifiers  
1..12411  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="plasmid pVR1012x/s containing HIV genes"

ORIGIN  
Query Match 80.3%; Score 1973.4; DB 6; Length 12411;  
Best Local Similarity 90.9%; Pred. No. 6e-198;  
Matches 2176; Conservative 0; Mismatches 191; Indels 28; Gaps 6;  
QY 14 TGGCCGAGCCATGAGCCAGG---CCACGAGCGCAACATCTGATGCGAGCGGCAACT 70  
Db 2967 TGGCCGAGCCATGAGCCAGGTCACACAGCGCCACCATCATGATGAGCGCGGCAACT 3026  
QY 71 TCAAGGGCCCCAAGCGCATCATCAAGTGTTCATCTGCGCAAGGAGGCGGCACATCGCCC 130  
Db 3027 TCGCAACACGAGCGCAAGATCGTGAAGTGTCTCAACTGGCGCAAGGAGGCGGCACACCGCCC 3086

QY 131 GCAACTGCGCGCCCGCCCGCAAGAGGCTGCTGGAAATGCGGCAAGGAGGCGCAACAGA 190  
Db 3087 GCAACTGCGCGCCCGCCCGCAAGAGGCTGCTGGAAATGCGGCAAGGAGGCGCAACAGA 3146  
QY 191 TGAAGGACTTGCACCGAGCGCCAGGCAACTTCTTCCCGAGGAGCTTGGCTTCCCCAGG 250  
Db 3147 TGAAGGACTTGCACCGAGCGCAAGGCTTA-TTTTATAGGAGATCTGGCTTCCCAAG 3205  
QY 251 GCAAGGCCGCGAGTTCGCCAGCGAGAGAACCGGCGCAACAGCCCCACAGCGCGGAGC 310  
Db 3206 GGAAGGCCGAGGAAATTTTCTTCAAGAGCAGACCAAGAGCAACAGCCCCACAGAGAGAGC 3265  
QY 311 TGCAGGTGCGCG-----CGACAACCCCGCCAGAGGCGCGCGCGAGCGCGCAGGGCA 364  
Db 3266 TTCAGGTTTGGGAGAGACACACTCCCTCTCAGAGCAGAGCGCGATAGACNAGGA 3325  
QY 365 -----CCCTGAACTTCCCGCCAGATCACTGTGTGAGCGCGCCCTGTGTGAGCATCAAG 418  
Db 3326 CTGTATCTTCTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCTCTGTCACAATAAGA 3385  
QY 419 TGGCGCGGCAGATCAAGAGGCGCTGTGGACACCGCGCGCGAGCAGACACCGTCTGAGG 478  
Db 3386 TAGGGGGCCAGCTGAAGGAGGCGCTTCTAGACACCGCGCGCGAGCAGACCGTCTGAGG 3445  
QY 479 AGATGAGCTGCGCGCGAGTGGAGGCGCAAGATGATCGCGCGCATCGCGCGGTTCATCA 538  
Db 3446 AGATCAACCTGCGCGCGCGCTGGAAGCGCAAGATGATCGCGCGCATCGCGCGGTTCATCA 3505  
QY 539 AGGTGCGCAGTACGACCGAGATCTCTGATCGAGATCTCGCGCAAGAGGCCATCGGACCG 598  
Db 3506 AGGTGCGCAGTACGACCGAGATCTCTGATCGAGATCTCGCGCGCAAGAGGCCATCGGACCG 3565  
QY 599 TGCTGATCGCT 658  
Db 3566 TGCTGATCGCT 3625  
QY 659 GCACCTCGAACTTCCCATCAGCGCCCATCGAGACCGCTGCGCGCGCGCGCGCGCGCGCGCA 718  
Db 3626 GCACCTCGAACTTCCCATCAGCGCCCATCGAGACCGCTGCGCGCGCGCGCGCGCGCGCGCA 3685  
QY 719 TGGACGCT 778  
Db 3686 TGGACGCT 3745  
QY 779 CCATCTGCGAGGAGTGGAGAGGAGGCGCAAGATCAACAGATCGCGCGCGCGCGCGCGCGCGCT 838  
Db 3746 AGATCTGCGAGGAGTGGAGAGGAGGCGCAAGATCAACAGATCGCGCGCGCGCGCGCGCGCGCT 3805  
QY 839 ACAACACCGCGCTGTTCGCGCATCAAGAGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 898  
Db 3806 ACAACACCGCGCTGTTCGCGCATCAAGAGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3865  
QY 899 ACTTCCGCGAGCTGAAACAAAGCGCACCGAGACTTCTGGAGGTCAGAGCTGGGCGATCCCCC 958  
Db 3866 ACTTCCGCGAGCTGAAACAAAGCGCACCGAGACTTCTGGAGGTCAGAGCTGGGCGATCCCCC 3925  
QY 959 ACCCGCGCGCGCTGAAAGAGAGAGCGTGTGCGCTGTGAGACGTGGGCGAGCGCTACT 1018  
Db 3926 ACCCGCGCGCGCTGAAAGAGAGAGCGTGTGCGCTGTGAGACGTGGGCGAGCGCTACT 3985  
QY 1019 TCAGCGTGCCTCTGAGAGGAGCTTCCGCAAGTACACCGCTTTCACCATCTCCCGAGCATCA 1078  
Db 3986 TCAGCGTGCCTCTGAGAGGAGCTTCCGCAAGTACACCGCTTTCACCATCTCCCGAGCATCA 4045  
QY 1079 ACAACGAGACCGCGCGCTGCGCTTACAGTACACGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCA 1138  
Db 4046 ACAACGAGACCGCGCGCTGCGCTTACAGTACACGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCA 4105  
QY 1139 GCCCGAGCATCTTCAGAGCAGCATGACCAAGATCTCTGAGCGCGCTTCCGCGCGCGCGCAAC 1198  
Db 4106 GCCCGAGCATCTTCAGTGCAGCATGACCAAGATCTCTGAGCGCGCTTCCGCGCGCGCGCAAC 4165

QY	1199	CCGAGATCGTGATCTTACCA-----GGCCCCCTGTAGTGGGCAGCGACTCGAGATCG	1252
DB	4166	CCGACATCGTGATCTACCAAGTACATGGGACGACTGTAGTGGGAGCGACTCGGAGATCG	4225
QY	1253	GCAGACCGCGCCAAAGATCGAGGAGCTGCGCAAGCACCTCTCGCTGGGGCTTCACCA	1312
DB	4226	GCAGACCGCACCAAGATCGAGGAGCTGCGCACACCTCTCGCTGGGGCTTCACCA	4285
QY	1313	CCCCCGACAAGAAGCAACAGAGGAGCCCCCTTCTGCCCCAT-----CGAGCTGCAAC	1366
DB	4286	CCCCCGACAAGAAGCAACAGAGGAGCCCCCTTCTTGTGATGGGTACGAGCTGCACC	4345
QY	1367	CCGACAAGTGGACCGGTGACGCCCATCGAGCTGCCGAGAAGGAGGTGGACCTGAAAG	1426
DB	4346	CCGACAAGTGGACCGGTGACGCCCATCGTCTGCCGAGAAGGACAGTGGACCGTGAACG	4405
QY	1427	ACATCCAGAAGCTGGTGGCGAGCTGAATCTGGCGGCAGCCAGATCTACCCCCGGCATCAAG	1486
DB	4406	ACATCCAGAAGCTGGTGGCGAGCTGAATCTGGCGGCAGCCAGATCTACCCCCGGCATCAAG	4465
QY	1487	TGGCCAGCTGTGTCAAGACTGCTGCGCGGGGCCAAAGCCCTGTGACCCGACATCGTGCCTCTGA	1546
DB	4466	TGGCCAGCTGTGTCAAGACTGCTGCGCGGCACCAAGGCCCTGTGACCCGAGTGTGTGCCTCTGA	4525
QY	1547	CCGAGGAGCGGAGCTGTGAGTGTGCCAGAAACCGCGGAGTCTCTGCGGAGCCCGTGGACG	1606
DB	4526	CCGAGGAGCGGAGCTGTGAGTGTGCCAGAAACCGCGAGATCTGTGAAGAGGCCCGTGGACG	4585
QY	1607	GCGTGTACTACGACCCCGACAGAGACTGCTGTGCGCCGAGATCCAGAAAGCAGGCCCACGACC	1666
DB	4586	GCGTGTACTACGACCCCGACAGAGACTGATCGCCGAGATCCAGAAAGCAGGCCCACGAGGCC	4645
QY	1667	AGTGGACCTTACAGATCTTACAGAGCGCTTCAAGAACCTGAAGACCGGACCGGAGTACGCCA	1726
DB	4646	AGTGGACCTTACAGATCTTACAGAGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCCCC	4705
QY	1727	AGATGCGCACCGGCCCAACCAAGAGAGTGTGAAGCAGTGTGACCCGAGGCGCTGTCAAGAAGATCG	1786
DB	4706	GCATGAAGGGCGGCCCAACCAACGAGTGTGAAGCAGTGTGACCCGAGGCGCTGTCAAGAAGATCG	4765
QY	1787	CCATGGAGAGCATCTGTGATCTGGGGCAGACCCCAAGTTCGCGCTGCCCATCCAGAAGG	1846
DB	4766	CCACCGAGAGCATCTGTGATCTGGGGCAGACCCCAAGTTCGAGTGTGCCCATCCAGAAGG	4825
QY	1847	AGACTGGGAGACCTGTGTGACCGACTACTTGGCAGGCCACCTGGATCCCCGAGTGGGAGT	1906
DB	4826	AGACTGGGAGGCTGTGTGACCGAGTACTTGGCAGGCCACCTGGATCCCCGAGTGGGAGT	4885
QY	1907	TCGTGAACACCCCCCTCTGTGAAGCTGTGTGATCAGAGTGGAGAGAGGCCCATCATCG	1966
DB	4886	TCGTGAACACCCCCCTCTGTGAAGCTGTGTGATCAGAGTGGAGAGAGGCCCATCATCG	4945
QY	1967	CGCGCGAGACCTTCTAGTGGACCGCGCCGCAACCGCGAGACAAGATCGGCAAGGCGCG	2026
DB	4946	CGCGCGAGACCTTCTAGTGGACCGCGCGCGCAACCGCGAGACAAGTGGGCAAGGCGCG	5005
QY	2027	GCTAGTACCGACGACCGGGCGCGGAGAGATCTGTGAGCCTGTGACCGAGACCAACCAACGAG	2086
DB	5006	GCTAGTACCGACGCGCGGCCCGCAGAAAGTGTGTGCCCTGTGACCGACCAACCAACGAG	5065
QY	2087	AGACCGAGCTGAGGCCATCCAGTGTGCCCTGTGAGGACAGCGGACCGAGGTGAACATCG	2146
DB	5066	AGACCGAGCTGAGGCCATCCACTTGGCCCTGTGAGGACAGCGGCTGTGAGGTGAACATCG	5125
QY	2147	TGACCGCACGCGCAGTACGCCCTTGGGCATCTCCAGGCCACGCGCACGAGCGAGAGCG	2206
DB	5126	TGACCGCACGCGCAGTACGCCCTTGGGCATCTCCAGGCCACGCGCACGAGCGAGAGCG	5185
QY	2207	AGCTGGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAAAGTGTACTGTAGCTGGG	2266
DB	5186	AGCTGGTGAAGCAGATCATCGAGCAGCTGATCAAGAGGAGAAAGTGTACTGTGGCTGGG	5245
QY	2267	TGCCCGCCCAAGGGCATTCGCGGCACACGAGCAGATCGACACGCTGGTGAACAGGGCA	2326

Sequence 12, Appl  
Sequence 48, Appl  
Sequence 42, Appl  
Sequence 46, Appl  
Sequence 40, Appl  
Sequence 47, Appl  
Sequence 41, Appl  
Sequence 30, Appl  
Sequence 33, Appl  
Sequence 32, Appl  
Sequence 13, Appl  
Sequence 14, Appl  
Sequence 13, Appl  
Sequence 17, Appl  
Sequence 14, Appl  
Sequence 13, Appl  
Sequence 9, Appl  
Sequence 9, Appl  
Sequence 10, Appl  
Sequence 11, Appl  
Sequence 10, Appl  
Sequence 11, Appl  
Sequence 54, Appl  
Sequence 54, Appl  
Sequence 75, Appl  
Sequence 59, Appl  
Sequence 83, Appl  
Sequence 16, Appl  
Sequence 31, Appl  
Sequence 31, Appl

US-10-190-305A-12  
US-10-190-305A-42  
US-10-190-305A-48  
US-10-190-305A-42  
US-10-190-305A-46  
US-10-190-305A-40  
US-10-190-305A-47  
US-10-190-305A-41  
US-10-190-305A-30  
US-10-190-305A-32  
US-10-241-009-13  
US-10-241-009-13  
US-10-241-009-14  
US-10-190-305A-13  
US-10-190-305A-17  
US-10-190-305A-13  
US-10-241-009-9  
US-10-190-305A-9  
US-10-241-009-10  
US-10-241-009-11  
US-10-190-305A-10  
US-10-190-305A-11  
US-10-241-009-54  
US-10-190-305A-54  
US-10-190-305A-55  
US-10-190-305A-59  
US-10-190-305A-83  
US-10-190-305A-16  
US-10-190-305A-16  
US-10-241-009-31  
US-10-190-305A-31

## ALIGNMENTS

```

1  RESULT 1
2  US-10-190-435-43
3  / Sequence 43, Application US/10190435
4  / Publication No. US20030143248A1
5  / GENERAL INFORMATION:
6  / APPLICANT: ZUR MEGEDE, Jan
7  / APPLICANT: BARNETT, Susan W.
8  / APPLICANT: LIAN, Ying
9  / APPLICANT: ENGELBRECHT, Susan
10 / APPLICANT: VAN RENSBURG, Estrellita
11 / TITLE OF INVENTION: POLYNUCLEOTIDES
12 / TITLE OF INVENTION: POLYPEPTIDES, 1
13 / FILE REFERENCE: PP18133.003 / 2302-1
14 / CURRENT APPLICATION NUMBER: US/10/1-
15 / CURRENT FILING DATE: 2002-12-30
16 / NUMBER OF SEQ ID NOS: 319
17 / SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ ID NO 43
; LENGTH: 2445
; TYPE: DNA
; ORGANISM: Artibeus
; FEATURE:
; OTHER INFORMATION:
US-10-190-435-43

```

```

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2pol.opt.YMMW_C
US-10-190-435-43

```

	Query Match	99.4%	Score 2443.4	DB 14	Length 2445
	Best Local Similarity	100.0%	Pred. No. 0		
	Matches 2444	Conservative	0	Mismatches 1	Indels 0
					Gaps 0
QY	7	GCACCATGCGCAGGCCATGAGCCAGGCCACAGGCCCAACATCTCTGATGCAGCGCAGC	66		
Db	1	GCACCATGCGCAGGCCATGAGCCAGGCCACAGGCCCAACATCTCTGATGCAGCGCAGC	60		
QY	67	AACATTCAAGGGCCCCAAGCGCATCATCAAGTGCCTTCAACTGCGGCAAGAGGGCCACATC	126		
Db	61	AACATTCAAGGGCCCCAAGCGCATCATCAAGTGCCTTCAACTGCGGCAAGAGGGCCACATC	120		
QY	127	GCCGCGCAACTGCGCGCGCCCCCGCCGACAGAGGGCTGCTGGAATGGCGCAAGAGGGCCAC	186		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	2443.4	99.4	2445	14	US-10-190-435-43	Sequence 37, Appl	
2	2443.4	99.4	2445	14	US-10-190-305A-37	Sequence 9, Appl	
3	2436.4	99.2	3930	14	US-10-190-435-9	Sequence 10, Appl	
4	2434.8	99.1	3930	14	US-10-190-435-10	Sequence 11, Appl	
5	2434.8	99.1	3930	14	US-10-190-435-11	Sequence 58, Appl	
6	2434.8	99.1	5184	14	US-10-190-435-58	Sequence 82, Appl	
7	2434.8	99.1	5184	14	US-10-190-305A-82	Sequence 44, Appl	
8	2428.6	98.8	2457	14	US-10-190-435-44	Sequence 38, Appl	
9	2428.6	98.8	2457	14	US-10-190-305A-38	Sequence 13, Appl	
10	2404.4	97.9	3531	14	US-10-190-435-13	Sequence 45, Appl	
11	2403.4	97.8	2457	14	US-10-190-435-15	Sequence 39, Appl	
12	2403.4	97.8	2457	14	US-10-190-305A-39	Sequence 14, Appl	
13	2402.8	97.8	3537	14	US-10-190-435-14	Sequence 12, Appl	
14	2402.8	97.8	3537	14	US-10-190-435-15	Sequence 15, Appl	
15	2401.8	97.8	5145	14	US-10-190-435-12	Sequence 42, Appl	







914	ACAAGCGACCCAGGACTTCTGGAGGTGCACTGGGCATCCCCACACCCCGCGGCTGA	973
Qy		
2387	ACAAGCGCACCCAGGACTTCTGGAGGTGCACTGGGCATCCCCACACCCCGCGGCTGA	2446
Db		
974	AGAGAGAAGAGCGGTGACCGTGTCTGGACGTGGGCGACGCTTACTTCAAGCGTGCCTCTGG	1033
Qy		
2447	AGAGAGAAGAGCGGTGACCGTGTCTGGACGTGGGCGACGCTTACTTCAAGCGTGCCTCTGG	2506
Db		
1034	ACGAGGACTTCCGCAAGTACACCGCTTCAACATCCCAGCAGTCAACAAAGAGAGACCCCG	1093
Qy		
2507	ACGAGGACTTCCGCAAGTACACCGCTTCAACATCCCAGCAGTCAACAAAGAGAGACCCCG	2566
Db		
1094	GCATCCGCTACCAAGTACAAAGTGTCTCCCAAGGGTGGAAAGGGCAGCCCAAGCATCTTCC	1153
Qy		
2567	GCATCCGCTACCAAGTACAAAGTGTCTCCCAAGGGTGGAAAGGGCAGCCCAAGCATCTTCC	2626
Db		
1154	AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCTGTGATCT	1213
Qy		
2627	AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCTGTGATCT	2686
Db		
1214	ACCAGGCCCCCTGTACTGTTGGCAGCGACCTCGAGATCGGCAGACACCGCGCCAGATCG	1273
Qy		
2687	ACCAGGCCCCCTGTACTGTTGGCAGCGACCTCGAGATCGGCAGACACCGCGCCAGATCG	2746
Db		
1274	AGGAGCTGGCGAAGCAACCTGCTGCGCTGGGGTTTCAACCAACCCCGCAAAAGAACCCAGA	1333
Qy		
2747	AGGAGCTGGCGAAGCAACCTGCTGCGCTGGGGTTTCAACCAACCCCGCAAAAGAACCCAGA	2806
Db		
1334	AGGAGCCCCCTTCTTGCCCATCTGAGCTGCAACCCGACAAAGTGGACGCTGCAGCCCATCG	1393
Qy		
2807	AGGAGCCCCCTTCTTGCCCATCTGAGCTGCAACCCGACAAAGTGGACGCTGCAGCCCATCG	2866
Db		
1394	AGCTGCCGAGAGGAGAGCTGGACCGTGAACACATCCAGAAAGCTGTGGGCAAGCTGA	1453
Qy		
2867	AGCTGCCGAGAGGAGAGCTGGACCGTGAACACATCCAGAAAGCTGTGGGCAAGCTGA	2926
Db		
1454	ACTGGCGCAGCAGATCTTACCCCGGATCAAGGTGCGCAGCTGTGCAAGCTGCTCGCG	1513
Qy		
2927	ACTGGCGCAGCAGATCTTACCCCGGATCAAGGTGCGCAGCTGTGCAAGCTGCTCGCG	2986
Db		
1514	GGCCAAAGGCCCTGACCGCATCGTGCCTGTACCGAGAGGCGGAGCTGGAGCTGGCCG	1573
Qy		
2987	GGCCAAAGGCCCTGACCGCATCGTGCCTGTACCGAGAGGCGGAGCTGGAGCTGGCCG	3046
Db		
1574	AGAACCCGAGATCTCTGCGAGCCCGTGCACCGCGTGTACTACGACCCCGAGAGACC	1633
Qy		
3047	AGAACCCGAGATCTCTGCGAGCCCGTGCACCGCGTGTACTACGACCCCGAGAGACC	3106
Db		
1634	TGCTGGCGAGATCCAGAACAGGGCCAACACAGTGGACTTACCAGATCTTACAGAGC	1693
Qy		
3107	TGCTGGCGAGATCCAGAACAGGGCCAACACAGTGGACTTACCAGATCTTACAGAGC	3166
Db		
1694	CTTTCAAGAACCTTGAAACCGCGAAGTACGCAAGATGCGCACCGCCCAACCAACGACG	1753
Qy		
3167	CTTTCAAGAACCTTGAAACCGCGAAGTACGCAAGATGCGCACCGCCCAACCAACGACG	3226
Db		
1754	TGAAGCAGCTGAACGAGCGCGTGCAGAGATCGCCATGAGAGCATCGTGATCTGGGGCA	1813
Qy		
3227	TGAAGCAGCTGAACGAGCGCGTGCAGAGATCGCCATGAGAGCATCGTGATCTGGGGCA	3286
Db		
1814	AGACCCCAAGTTCGCGCTGCGCATCCAGAAGGAGACTGGGAGACTGTGTGAGCCGACT	1873
Qy		
3287	AGACCCCAAGTTCGCGCTGCGCATCCAGAAGGAGACTGGGAGACTGTGTGAGCCGACT	3346
Db		
1874	ACTGGCAGGCCACTGTGATCCCCGAGTGGGATTCGTGAACACCCCCCCCCCTGTGTGAAGC	1933
Qy		
3347	ACTGGCAGGCCACTGTGATCCCCGAGTGGGATTCGTGAACACCCCCCCCCCTGTGTGAAGC	3406
Db		
1934	TGTGGTACCAAGTTCGCGCTGCGCATCCAGAAGGAGCCATCTTACTGTGACCGCG	1993
Qy		
3407	TGTGGTACCAAGTTCGCGCTGCGCATCCAGAAGGAGCCATCTTACTGTGACCGCGCG	3466
Db		
1994	CCGCGAACCCGCGAGACCAAGATCGCGAAGCGCGGCTACGTGACCGACCGGGCCGCGAGA	2053
Qy		

## RESULT 4

US-10-190-435-10

US-10-190-433-10  
; Sequence 10, Application US/10190435  
: Publication No. US20030143248A1

Publication No. US20030143248A1

1. PUBLICATION NO: 0029  
2. GENERAL INFORMATION:

: APPLICANT: ZUR MEGEDE, Jan

APPLICANT: ZUK MEGEDE, VAN  
BARNETT, Susan W

APPLICANT:

APPLICANT: LIAN, LING  
APPLICANT: ENGELBRECHT, Susan

; APPLICANT: ENGELBRECHT, SUBBI  
: APPLICANT: VAN BENSURG. Estrelita J.

APPLICANTI: VAN KENSBURG, ESTHERICA O.  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C

;  
; TITLE OF INVENTION: POLYPEPTIDES ENCODING ANTIGENS AND  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

1. TITLE OF INVENTION: POLYMERIDES, FOLIE  
2. FILE REFERENCE: PP18133 003 / 2302-18133

FILE REFERENCE: PP18133.003 / 2302-18133  
CURRENT APPLICATION NUMBER: US/10/190.435

; CURRENT APPLICATION NUMBER: US/1  
 ; CURRENT FILING DATE: 2003-12-30

2002-12-30

; NUMBER OF SEQ

```

; SOFTWARE: Pat

```

; SEQ ID NO 10

; LENGTH: 3930

```
; TYPE: DN
```

; ORGANISM: Artificial Sequence

```

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagCompPolmutAtt_
US-10-190-435-10

```

[illegible]

Query Match	Score	Score
99.1%	2434.8	2434.8

Best Local Similarity 99.9%; Pred. No. 0; 0: Gaps

Matches	2436;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;
---------	-------	--------------	----	------------	----	--------	----	------	----

-----TTCAGCGGCACTTCA 73

14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAAACATCTGATGCAGCGCAGCAACTCA / 3

1546

b  
1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAAACATCCTGATGCAGCGCAGCAACTTCA 1549

133



QY 194 AGGACTGACCGAGCGCCAGGCGAACTTCTTCGCGAGGACCTGCGCTTCCCGACGGCA 253  
Db 1667 AGGACTGACCGAGCGCCAGGCGAACTTCTTCGCGAGGACCTGCGCTTCCCGACGGCA 1726  
QY 254 AGGCGCGAGTTCCTCCAGCAGAGAACCGCGCCAAACAGCCCGCCACCGAGCGCGAGCTGC 313  
Db 1727 AGGCGCGAGTTCCTCCAGCAGAGAACCGCGCCAAACAGCCCGCCACCGAGCGCGAGCTGC 1786  
QY 314 AGGTGCGCGGAGCAAAACCCCGAGAGAGGCGCGCGCGCGAGCGCGAGGCGACCTGGAACCT 373  
Db 1787 AGGTGCGCGGAGCAAAACCCCGAGAGAGGCGCGCGCGCGAGCGCGAGGCGACCTGGAACCT 1846  
QY 374 TCCCGCAGATCACCTGTGCGAGCGCCCTGTGTGAGCATCAAGGTGGCGCGCGAGATCA 433  
Db 1847 TCCCGCAGATCACCTGTGCGAGCGCCCTGTGTGAGCATCAAGGTGGCGCGCGAGATCA 1906  
QY 434 AGGAGCCCTGTGTGAGACACCGCGCGCGAGACACCGTGTGTGAGAGATGAGCTGCGCG 493  
Db 1907 AGGAGCCCTGTGTGAGACACCGCGCGCGAGACACCGTGTGTGAGAGATGAGCTGCGCG 1966  
QY 494 GCAAGTGGAGCCCAAGATGATCGCGGATCGCGGATCGCGGCTTTCATCAAGGTGGCGCGAGTAGC 553  
Db 1967 GCAAGTGGAGCCCAAGATGATCGCGGATCGCGGATCGCGGCTTTCATCAAGGTGGCGCGAGTAGC 2026  
QY 554 ACCAGATCCTGTATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGTGTATCGGCCCA 613  
Db 2027 ACCAGATCCTGTATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGTGTATCGGCCCA 2086  
QY 614 CCCCGGTGAATCATCTGCGCGCAACATGCTGACCCAGCTGGCTGCGCTGCACTTCC 673  
Db 2087 CCCCGGTGAATCATCTGCGCGCAACATGCTGACCCAGCTGGCTGCGCTGCACTTCC 2146  
QY 674 CCATCAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCATGGACGCGCCCAAGG 733  
Db 2147 CCATCAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCATGGACGCGCCCAAGG 2206  
QY 734 TGAAGCATGTGCCCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTCGAGGAGA 793  
Db 2207 TGAAGCATGTGCCCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTCGAGGAGA 2266  
QY 794 TGGAGAAGGAGGCAAGATCACAAGATCGGCCCGCGAGAACCCCTACAAACCCCCGTGT 853  
Db 2267 TGGAGAAGGAGGCAAGATCACAAGATCGGCCCGCGAGAACCCCTACAAACCCCCGTGT 2326  
QY 854 TCGGCATCAAGAGAAGGAGACGACCAAGTGGCGCAAGCTGGTGAGCTTCCGAGCTGA 913  
Db 2327 TCGGCATCAAGAGAAGGAGACGACCAAGTGGCGCAAGCTGGTGAGCTTCCGAGCTGA 2386  
QY 914 ACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCGCACCCCGCGGCTGA 973  
Db 2387 ACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCGCACCCCGCGGCTGA 2446  
QY 974 AGAAGAAGAGAGCGTGAACCGTGTGAACGTGGGCGACGCTTACTTCAGCGTGCCTGG 1033  
Db 2447 AGAAGAAGAGAGCGTGAACCGTGTGAACGTGGGCGACGCTTACTTCAGCGTGCCTGG 2506  
QY 1034 ACGAGGACTCCGCAAGTACACCGCTTACCATCCCGACATCAACAGAGACCCCG 1093  
Db 2507 ACGAGGACTCCGCAAGTACACCGCTTACCATCCCGACATCAACAGAGACCCCG 2566  
QY 1094 GCATCCGCTACCACTACAAGTGTGCTCCCGAGGCTGGAGAGGCGAGCCCGATTTTC 1153  
Db 2567 GCATCCGCTACCACTACAAGTGTGCTCCCGAGGCTGGAGAGGCGAGCCCGATTTTC 2626  
QY 1154 AGAGCAGATGACCAAGATCTGAGCGCTTCCGCGCCGCAACCCCGAGATCGTGATCT 1213  
Db 2627 AGAGCAGATGACCAAGATCTGAGCGCTTCCGCGCCGCAACCCCGAGATCGTGATCT 2686  
QY 1214 ACCAGGCCCTCTGTGTGCGGACGACCTTGGAGATCGGCCAGACCGCGCCAAAGTCG 1273  
Db 2687 ACCAGGCCCTCTGTGTGCGGACGACCTTGGAGATCGGCCAGACCGCGCCAAAGTCG 2746  
QY 1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGCTTCAACACCCCGCAAGAGCACCGA 1333

Db 2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGCTTCAACACCCCGCAAGAGCACCGA 2806  
QY 1334 AGAGCCCCCTTCTTCCCATCTGAGCTGCACCCCGAAGTGGACCTGCGAGCCCATCG 1393  
Db 2807 AGAGCCCCCTTCTTCCCATCTGAGCTGCACCCCGAAGTGGACCTGCGAGCCCATCG 2866  
QY 1394 AGCTGCCCGAAGAGAGAGCTGGAACCGTGAACGACATCCAGAAAGCTGTGGGCAAGCTGA 1453  
Db 2867 AGCTGCCCGAAGAGAGAGCTGGAACCGTGAACGACATCCAGAAAGCTGTGGGCAAGCTGA 2926  
QY 1454 ACTGGCGAGCCAGATCTACCCCGGATCAAGGTGGCGCAGCTGTGTCAAGCTGTGCGG 1513  
Db 2927 ACTGGCGAGCCAGATCTACCCCGGATCAAGGTGGCGCAGCTGTGTCAAGCTGTGCGG 2986  
QY 1514 GCGCAAGGCCCTGACCGACATCGTCCCTGACCGAGAGGCGGAGCTGGAGCTGCGCG 1573  
Db 2987 GCGCAAGGCCCTGACCGACATCGTCCCTGACCGAGAGGCGGAGCTGGAGCTGCGCG 3046  
QY 1574 AGAACCGCGAGATCTTGGCGGAGCCCGTGCACCGCGCTGTACTACGACCCCGAGAGACC 1533  
Db 3047 AGAACCGCGAGATCTTGGCGGAGCCCGTGCACCGCGCTGTACTACGACCCCGAGAGACC 3106  
QY 1634 TGGTGGCGGAGATCCAGAGCAGGCCACGACAGTGGACCTACAGATCTACAGAGG 1693  
Db 3107 TGGTGGCGGAGATCCAGAGCAGGCCACGACAGTGGACCTACAGATCTACAGAGG 3166  
QY 1694 CTTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAGATGCGCACCGGCCACACCAAGACG 1753  
Db 3167 CTTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAGATGCGCACCGGCCACACCAAGACG 3226  
QY 1754 TGAAGCAGCTGACCGAGGCGCTGAGAGATCGCCATGGAGAGCATCGTATCTGGGGCA 1813  
Db 3227 TGAAGCAGCTGACCGAGGCGCTGAGAGATCGCCATGGAGAGCATCGTATCTGGGGCA 3286  
QY 1814 AGACCCCAAGTTCGCGCTGCCATCCAGAAAGAGACCTTGGGAGACCTGTGTGACCGACT 1873  
Db 3287 AGACCCCAAGTTCGCGCTGCCATCCAGAAAGAGACCTTGGGAGACCTGTGTGACCGACT 3346  
QY 1874 ACTGGAGGCCACCTGATCCCGAGTGGAGTCTGTGAACACCCCGCTTGGTGAAGC 1933  
Db 3347 ACTGGAGGCCACCTGATCCCGAGTGGAGTCTGTGAACACCCCGCTTGGTGAAGC 3406  
QY 1934 TGTGTGTCACAGCTGGAGAGGAGCCCATCATCGCGCGCGAGACCTTACGTGAGACGCG 1993  
Db 3407 TGTGTGTCACAGCTGGAGAGGAGCCCATCATCGCGCGCGAGACCTTACGTGAGACGCG 3466  
QY 1994 CCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAGCGGCGCGGAG 2053  
Db 3467 CCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAGCGGCGCGGAG 3526  
QY 2054 AGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGACGAGCCATCCAGCTGG 2113  
Db 3527 AGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGACGAGCCATCCAGCTGG 3586  
QY 2114 CCCTGAGGACAGCGGAGCGAGGTGAACATCGTGAACGAGCAGCAGTACGCTCGGCA 2173  
Db 3587 CCCTGAGGACAGCGGAGCGAGGTGAACATCGTGAACGAGCAGCAGTACGCTCGGCA 3646  
QY 2174 TCATCCAGGCCCGCGAGACGAGCGAGCTGTGTGAACAGATCATCGAGCAGC 2233  
Db 3647 TCATCCAGGCCCGCGAGACGAGCGAGCTGTGTGAACAGATCATCGAGCAGC 3706  
QY 2234 TGATCAAGAGGAGAGGTGTACTGTGCTGGTGCCTGCGGCCCAAGAGGATCGGCGCA 2293  
Db 3707 TGATCAAGAGGAGAGGTGTACTGTGCTGGTGCCTGCGGCCCAAGAGGATCGGCGCA 3766  
QY 2294 ACGAGCAGATCGACAGTGGTGGAGGAGGCGCATCCGCAAGTGTCTGTCTGAGCAGCA 2353  
Db 3767 ACGAGCAGATCGACAGTGGTGGAGGAGGCGCATCCGCAAGTGTCTGTCTGAGCAGCA 3826  
QY 2354 TCGATGGCGCATCGTGATCTACAGTACATGGACGCTGTACTGTGGGAGCGCGGCGC 2413



Db 3827 TCAGTGGCGGCGTGTGATCTACAGTACATGACGACCTGTACGTGGCGAGCGGGGCC 3886

Qy 2414 CTAGGATCGATTAAAGCTTCCGGGGCTAGCACCGGT 2451

Db 3887 CTAGGATCGATTAAAGCTTCCGGGGCTAGCACCGGT 3924

RESULT 5

US-10-190-435-11

; Sequence 11, Application US/10190435

; Publication No. US20030143248A1

; GENERAL INFORMATION:

; APPLICANT: ZUR MEDEDE, Jan

; APPLICANT: BARNETT, Susan W.

; APPLICANT: LIAN, Ying

; APPLICANT: ENGELBRECHT, Susan

; APPLICANT: VAN RENSBURG, Estrelita J.

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C

; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

; FILE REFERENCE: P18133.003 / 2302-18133

; CURRENT APPLICATION NUMBER: US/10/190,435

; CURRENT FILING DATE: 2002-12-30

; NUMBER OF SEQ ID NOS: 319

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 11

; LENGTH: 3930

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutIna\_C

US-10-190-435-11

Query Match 99.1%; Score 2434.8; DB 14; Length 3930;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 14 TGGCCGAGGCGCATGAGCAGGCGCACAGCGCCCAACATCTCTGATGCGAGCGCAACTTCA 73

Db 1487 TCGCCGAGGCGCATGAGCAGGCGCACAGCGCCCAACATCTCTGATGCGAGCGCAACTTCA 1546

Qy 74 AGGCGCCCAAGCGCATCATCAAGTGCTTCAACTGCGCAAGGAGGGGCCACATCGCCCGCA 133

Db 1547 AGGCGCCCAAGCGCATCATCAAGTGCTTCAACTGCGCGCAAGGAGGGGCCCATCGCCGCA 1606

Qy 134 ACTGCGCGGCGCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGAGGGGCCACCATGA 193

Db 1607 ACTGCGCGGCGCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGAGGGGCCACCATGA 1666

Qy 194 AGGACTGCAAGCGCGCGAGCGCAACCGCGCCCAACAGCGCCCGCGAGCTGC 253

Db 1667 AGGACTGCAAGCGCGCGAGCGCAACCGCGCCCAACAGCGCCCGCGAGCTGC 1726

Qy 254 AGGCGCGCGAGTTCGCCGAGCGAGCAGAAACCGCGCCCAACAGCGCCCGCGAGCTGC 313

Db 1727 AGGCGCGCGAGTTCGCCGAGCGAGCAGAAACCGCGCCCAACAGCGCCCGCGAGCTGC 1786

Qy 314 AGGTGCGCGCGCAACACCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373

Db 1787 AGGTGCGCGCGCAACACCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846

Qy 374 TCCCGCGAGTCAACCGTGTGGCAGCGCGCCCTGTGTGAGCATCAAGTGGCGCGCGCGCG 433

Db 1847 TCCCGCGAGTCAACCGTGTGGCAGCGCGCCCTGTGTGAGCATCAAGTGGCGCGCGCGCG 1906

Qy 434 AGGAGGCGCTGTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493

Db 1907 AGGAGGCGCTGTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1966

Qy 494 GCAAGTGGAGCCCAAGTGTATCGCGCGCATCGCGCGCTTCATCAAGTGGCGCGCGCGCG 553

Db 1967 GCAAGTGGAGCCCAAGTGTATCGCGCGCATCGCGCGCTTCATCAAGTGGCGCGCGCGCG 2026

Qy 554 ACCAGATCCTGATGAGATCTGCGCGCAAGAGGCGCCATCGGCGACCGTGTGATCGGCGCCA 613

Db 2027 ACCAGATCTGTGATCGAGTCTGGGCDAGAGGCCATCGGACCGTGTGATCGGCGCCA 2086

Qy 614 CCCCCTGTGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGGCTGACCCCTGAACCTTC 673

Db 2087 CCCCCTGTGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGGCTGACCCCTGAACCTTC 2146

Qy 674 CCATCAGCCCATCGAGACCGTGCCTGAGCTGAAGCCCGGCGCATGGAAGCGGCGCCCAAGG 733

Db 2147 CCATCAGCCCATCGAGACCGTGCCTGAGCTGAAGCCCGGCGCATGGAAGCGGCGCCCAAGG 2206

Qy 734 TGAAGCAGTGGCCCTGACCCGAGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793

Db 2207 TGAAGCAGTGGCCCTGACCCGAGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy 794 TGGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGGAGAAACCCCTACAAACACCCCTGT 853

Db 2267 TGGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGGAGAAACCCCTACAAACACCCCTGT 2326

Qy 854 TCGCCATCAAGAAGAAGAGCAGACCAAGTGGGCGCAAGCTGTGTGAGATCTTCGCGAGCTGA 913

Db 2327 TCGCCATCAAGAAGAAGAGCAGACCAAGTGGGCGCAAGCTGTGTGAGATCTTCGCGAGCTGA 2386

Qy 914 ACAAGCGCACCCAGGATCTTCTGGAGGTGCGCATCGGCCATCCGCCACCCCGCGGCTGA 973

Db 2387 ACAAGCGCACCCAGGATCTTCTGGAGGTGCGCATCGGCCATCCGCCACCCCGCGGCTGA 2446

Qy 974 AGAAGAGAGAGCGGTGACCGTGTGACGCTGGGCGACGCTACTTCAAGGTGCCCTGG 1033

Db 2447 AGAAGAGAGAGCGGTGACCGTGTGACGCTGGGCGACGCTACTTCAAGGTGCCCTGG 2506

Qy 1034 ACGAGGATCTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAGAGAGCCCGCG 1093

Db 2507 ACGAGGATCTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAGAGAGCCCGCG 2566

Qy 1094 GCATCCGCTACAGTACAACTGTGCTGCCAGGCGCTGGAAGGCGAGCGCCAGCATCTTCC 1153

Db 2567 GCATCCGCTACAGTACAACTGTGCTGCCAGGCGCTGGAAGGCGAGCGCCAGCATCTTCC 2626

Qy 1154 AGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCGCGCAACCCCGAGATCTGTGATCT 1213

Db 2627 AGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCGCGCAACCCCGAGATCTGTGATCT 2686

Qy 1214 ACCAGGCGCCCTGTACGTGGGCGAGCGACCTGGAGATCGGCGAGCACCGCGCCCAAGATCG 1273

Db 2687 ACCAGGCGCCCTGTACGTGGGCGAGCGACCTGGAGATCGGCGAGCACCGCGCCCAAGATCG 2746

Qy 1274 AGGAGCTGCGCAAGCACCTGTGCGCTGGGCTTCAACACCCCGAGCAAGAGCACCGAGA 1333

Db 2747 AGGAGCTGCGCAAGCACCTGTGCGCTGGGCTTCAACACCCCGAGCAAGAGCACCGAGA 2806

Qy 1334 AGGAGCGCCCTTCTGCGCATCGAGTGCACCCGAGACCGTGCAGCGCCATCG 1393

Db 2807 AGGAGCGCCCTTCTGCGCATCGAGTGCACCCGAGACCGTGCAGCGCCATCG 2866

Qy 1394 AGCTGCCGAGAGAGAGAGTGGACCGTGAACGACATCCAGAGCTGTGTGGCGCAAGCTGA 1453

Db 2867 AGCTGCCGAGAGAGAGAGTGGACCGTGAACGACATCCAGAGCTGTGTGGCGCAAGCTGA 2926

Qy 1454 ACTGGCGCCAGCAGATCTTACCCCGCATCAAGGTGCGCGCATGTGTGAAGTGTGTGCGCG 1513

Db 2927 ACTGGCGCCAGCAGATCTTACCCCGCATCAAGGTGCGCGCATGTGTGAAGTGTGTGCGCG 2986

Qy 1514 GCGCCAGGCGCTGACCGACATCTGTGCGCCCTGACCGAGGAGCGCGAGCTGGGCTGGCGG 1573

Db 2987 GCGCCAGGCGCTGACCGACATCTGTGCGCCCTGACCGAGGAGCGCGAGCTGGGCTGGCGG 3046

Qy 1574 AGAACCGCGAGATCTTGGCGGAGCGCGTGCACGCGGTGTACTAGACCCCGAGCAAGGAGCC 1633

Db 3047 AGAACCGCGAGATCTTGGCGGAGCGCGTGCACGCGGTGTACTAGACCCCGAGCAAGGAGCC 3106

Qy 1634 TGGTGGCGGAGATCCAGAGAGGCGCGACGACGAGTGCAGCTTACAGATCTTACAGGAGCG 1693

Db 3107 TGTGTCGCCAGATCCAGAGCAGGCGCCACACAGTGGACCTACAGATCTACAGGAGC 3166  
Qy 1694 CTTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGGACCCGCCACACACGACG 1753  
Db 3167 CTTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGGACCCGCCACACACGACG 3226  
Qy 1754 TGAAGCAGCTGACCGGAGGCGGTGCAAGATGCCATGAGAGCATCGTGTCTGGGCA 1813  
Db 3227 TGAAGCAGCTGACCGGAGGCGGTGCAAGATGCCATGAGAGCATCGTGTCTGGGCA 3286  
Qy 1814 AGACCCCAAGTTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCT 1873  
Db 3287 AGACCCCAAGTTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCT 3346  
Qy 1874 ACTGGCAGGCCACCTGGATCCCGGAGTGGGAGTTGGTAAACACCCCGCTGGTGAAGC 1933  
Db 3347 ACTGGCAGGCCACCTGGATCCCGGAGTGGGAGTTGGTAAACACCCCGCTGGTGAAGC 3406  
Qy 1934 TGTGTTACAGCTGGAGAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGAAGCGG 1993  
Db 3407 TGTGTTACAGCTGGAGAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGAAGCGG 3466  
Qy 1994 CGGCCAACCGCAGAGCAAGATCGCAAGCGCGGCTACGTGACCGAGCCGCGGCGGCA 2053  
Db 3467 CGGCCAACCGCAGAGCAAGATCGCAAGCGCGGCTACGTGACCGAGCCGCGGCGGCA 3526  
Qy 2054 AGATCTGAGCTGACCGGAGACCAACCAAGAGACCGAGTGTGTAACAGATCATCGAGCAGC 2113  
Db 3527 AGATCTGAGCTGACCGGAGACCAACCAAGAGACCGAGTGTGTAACAGATCATCGAGCAGC 3586  
Qy 2114 CCTGCGAGGACGCGGAGGAGTGAACATGTGACCGACAGCCAGTACGCGCTGGGCA 2173  
Db 3587 CCTGCGAGGACGCGGAGGAGTGAACATGTGACCGACAGCCAGTACGCGCTGGGCA 3646  
Qy 2174 TCATCCAGGCCACCGCGGAGAGGAGTGTGTAACAGATCATCGAGCAGC 2233  
Db 3647 TCATCCAGGCCACCGCGGAGAGGAGTGTGTAACAGATCATCGAGCAGC 3706  
Qy 2234 TGAATCAAGAGGAGGAGTGTACCTGAGTGGTGGCGCCGCCACAGGCGCATCGGCGCA 2293  
Db 3707 TGAATCAAGAGGAGGAGTGTACCTGAGTGGTGGCGCCGCCACAGGCGCATCGGCGCA 3766  
Qy 2294 ACCAGCAGATCGCAAGCTGGTGAAGGAGGATCCGCAAGTGTGTTCTTGGAGCGCA 2353  
Db 3767 ACCAGCAGATCGCAAGCTGGTGAAGGAGGATCCGCAAGTGTGTTCTTGGAGCGCA 3826  
Qy 2354 TCGATGGCGGATCGGTGATCTACCATGATGAGACGACCTGTACGTGGCAGCGGCGCC 2413  
Db 3827 TCGATGGCGGATCGGTGATCTACCATGATGAGACGACCTGTACGTGGCAGCGGCGCC 3886  
Qy 2414 CTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2451  
Db 3887 CTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 3924

RESULT 6  
US-10-190-435-58  
; Sequence 58, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEDEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: PP18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 58

; LENGTH: 5184  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: TatRevNefagCpolIna C  
US-10-190-435-58  
Query Match 99.1%; Score 2434.8; DB 14; Length 5184;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 14 TGGCCGAGGCGCATGAGCCAGGCGCACAGCGCCAAATCTCTGATGCGAGCGCAGCACTTCA 73  
Db 2741 TCGCCGAGGCGCATGAGCCAGGCGCACAGCGCCAAATCTCTGATGCGAGCGCAGCACTTCA 2800  
Qy 74 AGGGCCCCAAGCCCATCATCAAGTGTCTTCAACTGTGCGGCAAGGAGGCGCCATCGCCCGCA 133  
Db 2801 AGGGCCCCAAGCCCATCATCAAGTGTCTTCAACTGTGCGGCAAGGAGGCGCCATCGCCCGCA 2860  
Qy 134 ACTGCGCGCGCCCCCGCAAGAGGGGTGCTGGAAGTTCGGCAAGGAGGCGCCACCATGCA 193  
Db 2861 ACTGCGCGCGCCCCCGCAAGAGGGGTGCTGGAAGTTCGGCAAGGAGGCGCCACCATGCA 2920  
Qy 194 AGGACTGCAACCGAGCGCGAGGCCAACTTCTTCCGCGAGGACCTTGGGCTTCCCGCCAGGCA 253  
Db 2921 AGGACTGCAACCGAGCGCGAGGCCAACTTCTTCCGCGAGGACCTTGGGCTTCCCGCCAGGCA 2980  
Qy 254 AGGCCCGCGAGTTTCCCGAGCGAGCAAAACCGCGCCAAACAGCCCGCCAGCGGAGCTGC 313  
Db 2981 AGGCCCGCGAGTTTCCCGAGCGAGCAAAACCGCGCCAAACAGCCCGCCAGCGGAGCTGC 3040  
Qy 314 AGGTGCGCGCGCAAAACCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
Db 3041 AGGTGCGCGCGCAAAACCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3100  
Qy 374 TCCCGCCAGATCACCTGTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 423  
Db 3101 TCCCGCCAGATCACCTGTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3160  
Qy 434 AGGAGGCGCTGTGGACACCG 493  
Db 3161 AGGAGGCGCTGTGGACACCG 3220  
Qy 494 GCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAGTGGCGCGCATAG 553  
Db 3221 GCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAGTGGCGCGCATAG 3280  
Qy 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCAAGGCGCGCGCGCGCGCG 613  
Db 3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCAAGGCGCGCGCGCGCGCG 3340  
Qy 614 CCCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTC 673  
Db 3341 CCCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTC 3400  
Qy 674 CCATCAGCCCATCGAGACCGTGCCTGTGAGCTGAGCTGAGCCCGCGCATGAGCGCGCGCG 733  
Db 3401 CCATCAGCCCATCGAGACCGTGCCTGTGAGCTGAGCTGAGCCCGCGCATGAGCGCGCGCG 3460  
Qy 734 TGAAGCAGTGGCGCCCTGACCGAGAGAGATCAAGCGCTTGAACCGCGCATCTGCGAGGAGA 793  
Db 3461 TGAAGCAGTGGCGCCCTGACCGAGAGAGATCAAGCGCTTGAACCGCGCATCTGCGAGGAGA 3520  
Qy 794 TGGAGAGGAGGCGCAAGATCACCAAGTTCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 853  
Db 3521 TGGAGAGGAGGCGCAAGATCACCAAGTTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3580  
Qy 854 TCGCCATCAAGAGAGGAGCAGCAACCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 913  
Db 3581 TCGCCATCAAGAGAGGAGCAGCAACCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 3640  
Qy 914 ACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGCATCCCGCCAGCGCGCGCGCTGA 973

Db 3641 ACAGCGCCACCCAGGACTTCTGGAGGTCAGCTGGGCATCCCCACCCCGCGGCTGA 3700  
Qy 974 AGAAGAGAGACGCTGACGCTGCTGGAAGCTGGGAGCGCTACTTTCAGGTCGCCCTGG 1033  
Db 3701 AGAAGAGAGAGCGTGAACGCTGCTGGAAGCTGGGAGCGCTACTTTCAGGTCGCCCTGG 3760  
Qy 1034 ACCGAGACTTCCGCAAGTACACCGCTTACACCATCCCGAGCATCAACAACGAGACCCCG 1093  
Db 3761 ACAGGACTTCCGCAAGTACACCGCTTACACCATCCCGAGCATCAACAACGAGACCCCG 3820  
Qy 1094 GCATCCGCTACCAAGTACAAAGTGTGCTGCTCCAGGGCTGGAAGGCGAGCCCGAGCATTTCC 1153  
Db 3821 GCATCCGCTACCAAGTACAAAGTGTGCTGCTCCAGGGCTGGAAGGCGAGCCCGAGCATTTCC 3880  
Qy 1154 AGAGCAGCATGACCAAGATCCTGAGCCCTTCCGCGCCCAACCCCGAGATCGTGATCT 1213  
Db 3881 AGAGCAGCATGACCAAGATCCTGAGCCCTTCCGCGCCCAACCCCGAGATCGTGATCT 3940  
Qy 1214 ACCAGGCCCCCTGTAGCTGGGAGCGACCTGGAGATCGCCAGCAGCCGCGCAAGATCG 1273  
Db 3941 ACCAGGCCCCCTGTAGCTGGGAGCGACCTGGAGATCGCCAGCAGCCGCGCAAGATCG 4000  
Qy 1274 AGAGCTGCGCAAGCAGCTGCTGCTGGGCTTACCAACCCCGAGCAAGACCAAG 1333  
Db 4001 AGAGCTGCGCAAGCAGCTGCTGCTGGGCTTACCAACCCCGAGCAAGACCAAG 4060  
Qy 1334 AGAGCTGCGCAAGCAGCTGCTGCTGGGCTTACCAACCCCGAGCAAGACCAAG 1393  
Db 4061 AGAGCTGCGCAAGCAGCTGCTGCTGGGCTTACCAACCCCGAGCAAGACCAAG 4120  
Qy 1394 AGCTGCCGAGAGAGAGCTGACCGTGAACGACATCCAGAGCTGTGGGCAAGCTGA 1453  
Db 4121 AGCTGCCGAGAGAGAGCTGACCGTGAACGACATCCAGAGCTGTGGGCAAGCTGA 4180  
Qy 1454 ACTGGGCGAGCAGATCTACCCCGCATCAAGGTGCGCAGCTGTCAAGCTCTGCGCG 1513  
Db 4181 ACTGGGCGAGCAGATCTACCCCGCATCAAGGTGCGCAGCTGTCAAGCTCTGCGCG 4240  
Qy 1514 GCGCAAGGCGCTGACGACATGTCGCTGACCGAGAGGCGCGAGCTGGAGCTGGCGG 1573  
Db 4241 GCGCAAGGCGCTGACGACATGTCGCTGACCGAGAGGCGCGAGCTGGAGCTGGCGG 4300  
Qy 1574 AGAACCGGAGATCTCGGCGAGCCGTCGACCGCGTGTACTAGACCCCGCAAGGAC 1633  
Db 4301 AGAACCGGAGATCTCGGCGAGCCGTCGACCGCGTGTACTAGACCCCGCAAGGAC 4360  
Qy 1634 TGTGGCGGAGATCCAGAGCAGGCGCCACGACAGTGGACCTACCGATCTACCGAGG 1693  
Db 4361 TGTGGCGGAGATCCAGAGCAGGCGCCACGACAGTGGACCTACCGATCTACCGAGG 4420  
Qy 1694 CTTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGCGCACCGGCCACCAACGAG 1753  
Db 4421 CTTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGCGCACCGGCCACCAACGAG 4480  
Qy 1754 TGRAGCAGCTGACGAGCGCTGAGAGATGCGCATGGAGAGCATCGTGTGGGCA 1813  
Db 4481 TGRAGCAGCTGACGAGCGCTGAGAGATGCGCATGGAGAGCATCGTGTGGGCA 4540  
Qy 1814 AGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACT 1873  
Db 4541 AGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACT 4600  
Qy 1874 ACTGGCAGGCGACTTGATCCCCAGTGGAGTTCTGTAACACCCCGCCCTGGTGAAG 1933  
Db 4601 ACTGGCAGGCGACTTGATCCCCAGTGGAGTTCTGTAACACCCCGCCCTGGTGAAG 4660  
Qy 1934 TGTGGTACCACTGGAAGAGGCCATCATCGGCGCGAGACCTTCTAGCTGGAAGCGG 1993  
Db 4661 TGTGGTACCACTGGAAGAGGCCATCATCGGCGCGAGACCTTCTAGCTGGAAGCGG 4720  
Qy 1994 CCGCAACCGGAGACCAAGTTCGGCAAGGCGGCTACGTGAACGAGCGGCGCGGAG 2053  
Db 4721 CCGCAACCGGAGACCAAGTTCGGCAAGGCGGCTACGTGAACGAGCGGCGCGGAG 4780

Qy 2054 AGATCGTGAGCTGACCGAGACCAACAGAGCCGAGCTGCAGGCCATCCAGCTGG 2113  
Db 4781 AGATCGTGAGCTGACCGAGACCAACAGAGCCGAGCTGCAGGCCATCCAGCTGG 4840  
Qy 2114 CCCTGCGAGGACAGCGGCGAGGAGTGAACATCGTGACCGAGCAGTACGCCCTGGGCA 2173  
Db 4841 CCCTGCGAGGACAGCGGCGAGGAGTGAACATCGTGACCGAGCAGTACGCCCTGGGCA 4900  
Qy 2174 TCATCCAGGCGCCAGCGGACGAGCGAGCGAGCTGTGAACAGATCATCGAGCAGC 2233  
Db 4901 TCATCCAGGCGCCAGCGGACGAGCGAGCGAGCTGTGAACAGATCATCGAGCAGC 4960  
Qy 2234 TCATCAAGAGGAGAGAGTGTACCTGAGCTGGGTCGCCGCCCAAGAGGCAATCGCGGCA 2293  
Db 4961 TCATCAAGAGGAGAGAGTGTACCTGAGCTGGGTCGCCGCCCAAGAGGCAATCGCGGCA 5020  
Qy 2294 ACGAGCAGATCGACAGAGTGTGAGCAAGGCAATCGGCAAGGTCGTCTCTGAGCAGGCA 2353  
Db 5021 ACGAGCAGATCGACAGAGTGTGAGCAAGGCAATCGGCAAGGTCGTCTCTGAGCAGGCA 5080  
Qy 2354 TCGATGGGCGCATCGTGATCTACCAAGTACATGAGCAGCTGTACGTGGGCAAGCGCGGCC 2413  
Db 5081 TCGATGGGCGCATCGTGATCTACCAAGTACATGAGCAGCTGTACGTGGGCAAGCGCGGCC 5140  
Qy 2414 CTAGATCGATTAAAGCTTCCGCGGCTAGCACCGGT 2451  
Db 5141 CTAGATCGATTAAAGCTTCCGCGGCTAGCACCGGT 5178

## RESULT 7

US-10-190-305A-82  
; Sequence 82, Application US/10190305A  
; Publication No. US20030198621A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan  
; APPLICANT: LIAN, Ying  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: 2302-18702 / 18702.002  
; CURRENT APPLICATION NUMBER: US/10/190.305A  
; CURRENT FILING DATE: 2002-07-05  
; NUMBER OF SEQ ID NOS: 93  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 82  
; LENGTH: 5184  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: TatRevNefgagCpola C  
US-10-190-305A-82

Query Match 99.1%; Score 2434.8; DB 14; Length 5184;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 14 TGGCGAGGCCATGAGCCAGGCCACAGGCCCAACATCTGTATGCGAGCGCAGCACTTCA 73  
Db 2741 TGGCGAGGCCATGAGCCAGGCCACAGGCCCAACATCTGTATGCGAGCGCAGCACTTCA 2800  
Qy 74 AGGGCCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGAGGGGCCACATCGCCCGCA 133  
Db 2801 AGGGCCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGAGGGGCCACATCGCCCGCA 2860  
Qy 134 ACTGCGCGCCCCCGCGCAAGAGGCTCTGGAAGTGGGCAAGAGGGGCCACCAAGATGA 193  
Db 2861 ACTGCGCGCCCCCGCGCAAGAGGCTCTGGAAGTGGGCAAGAGGGGCCACCAAGATGA 2920  
Qy 194 AGGACTGCAACCGAGCGCCAGGCCCACTTCTTCCGAGGAGCTGCGCTTCCCGAGGCA 253  
Db 2921 AGGACTGCAACCGAGCGCCAGGCCCACTTCTTCCGAGGAGCTGCGCTTCCCGAGGCA 2980



Db 5141 CTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGT 5178

RESULT 8  
US-10-190-435-44  
; Sequence 44, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEHDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBERG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: PFI8133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 44  
; LENGTH: 2457  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: p2Poliopt.YM\_C  
US-10-190-435-44

Query Match 98.8%; Score 2428.6; DB 14; Length 2457;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 2447; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

Qy 1 GTCGACGCCACCATGGCGGAGGCGATGAGCCAGCCACAGCGCCACACATCTGTATGTCAG 60  
Db 1 GTCGAGGCCACCATGGCGGAGGCGATGAGCCAGCCACAGCGCCACACATCTGTATGTCAG 60

Qy 61 CGCGACCACTTCAGAGGCGCCCAAGCGCATCATCAAGTCTTCACTCGCGCAGAGGGGC 120  
Db 61 CGCGACCACTTCAGAGGCGCCCAAGCGCATCATCAAGTCTTCACTCGCGCAGAGGGGC 120

Qy 121 CACATCGCCGCCCAACTGCGCGCGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAG 180  
Db 121 CACATCGCGCAACTGCGCGCGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAG 180

Qy 181 GGCACACAGATGAAGAGCTGACCGAGGCGCAGCCAACTTCTCCCGGAGGACCTGGCC 240  
Db 181 GGCACACAGATGAAGAGCTGACCGAGGCGCAGCCAACTTCTCCCGGAGGACCTGGCC 240

Qy 241 TTCCCGCAGGCGCCGAGTCCCGAGTCCCGAGCGAGACCGCGCCACAGCCCGCAC 300  
Db 241 TTCCCGCAGGCGCCGAGTCCCGAGTCCCGAGCGAGACCGCGCCACAGCCCGCAC 300

Qy 301 AGCCGCGAGCTGAGTGGCGGCGACAAACCCCGCAGCGAGGCGCGCGCGAGCGCCAG 360  
Db 301 AGCCGCGAGCTGAGTGGCGGCGACAAACCCCGCAGCGAGGCGCGCGCGAGCGCCAG 360

Qy 361 GGCACCTTGAATTTCCCGCAGATCACTCTGTGGCAGCGCCCGCTGTGTGAGCATCAAGTG 420  
Db 361 GGCACCTTGAATTTCCCGCAGATCACTCTGTGGCAGCGCCCGCTGTGTGAGCATCAAGTG 420

Qy 421 GCGCGCCAGATCAAGAGGCGCTGTGTGACACCGCGCGCGAGCACACCGTGTGGAGGAG 480  
Db 421 GCGCGCCAGATCAAGAGGCGCTGTGTGACACCGCGCGCGAGCACACCGTGTGGAGGAG 480

Qy 481 ATGAGCTTCCCGCGCAAGTGAAGCCCAAGATGATCGCGGGGATCGCGGGCTTATCAAG 540  
Db 481 ATGAGCTTCCCGCGCAAGTGAAGCCCAAGATGATCGCGGGGATCGCGGGCTTATCAAG 540

Qy 541 GTGCGCCAGTACGACAGATCTGATCGAGATCTCGGCAAGAGGCGCATCGGACCGTG 600  
Db 541 GTGCGCCAGTACGACAGATCTGATCGAGATCTCGGCAAGAGGCGCATCGGACCGTG 600

Qy 601 CTGATCGGCCCCACCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
Db 601 CTGATCGGCCCCACCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy 661 ACCCTGAATCTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGATG 720  
Db 661 ACCCTGAATCTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGATG 720

Qy 721 GACGCGCCCAAGGTGAAGAGTGGCCCTGACCGAGGAGAGATCAAGAGGCGCTGACGCC 780  
Db 721 GACGCGCCCAAGGTGAAGAGTGGCCCTGACCGAGGAGAGATCAAGAGGCGCTGACGCC 780

Qy 781 ATCTCGAGGAGATGAGAGGAGGCGGAGATCAAGATCGGCGCGGAGAGAGCCCTAC 840  
Db 781 ATCTCGAGGAGATGAGAGGAGGCGGAGATCAAGATCGGCGCGGAGAGAGCCCTAC 840

Qy 841 AACACCCCGGTGTTCCGCATCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900  
Db 841 AACACCCCGGTGTTCCGCATCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900

Qy 901 TTTCGCGAGTGAACAAAGCGCACCGGACCTTCTGGAGGTGAGCTGGGATCCGCCAC 960  
Db 901 TTTCGCGAGTGAACAAAGCGCACCGGACCTTCTGGAGGTGAGCTGGGATCCGCCAC 960

Qy 961 CCGCGCGGCTGAAGAGAGAGGAGTGAACGCTGCTGGAGGTGAGCTGGGAGGAGGAGGAG 1020  
Db 961 CCGCGCGGCTGAAGAGAGAGGAGTGAACGCTGCTGGAGGTGAGCTGGGAGGAGGAGGAG 1020

Qy 1021 AGCGTGCCTTCGAGAGGAGTTCGCGAAGTACACCGCTTTCACCATCCCGAGCATCAAC 1080  
Db 1021 AGCGTGCCTTCGAGAGGAGTTCGCGAAGTACACCGCTTTCACCATCCCGAGCATCAAC 1080

Qy 1081 AACGAGACCCCGGATCGCTACAGTACAGGCTGCTGCGGAGGAGGAGGAGGAGGAGGAG 1140  
Db 1081 AACGAGACCCCGGATCGCTACAGTACAGGCTGCTGCGGAGGAGGAGGAGGAGGAGGAG 1140

Qy 1141 CCGAGCATCTTCGAGAGGAGTTCGCGAAGTACACCGCTTTCACCATCCCGAGCATCAAC 1200  
Db 1141 CCGAGCATCTTCGAGAGGAGTTCGCGAAGTACACCGCTTTCACCATCCCGAGCATCAAC 1200

Qy 1201 GAGATCGTATCTACAGGCGCCCTGCTGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260  
Db 1201 GAGATCGTATCTACAGGCGCCCTGCTGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260

Qy 1261 CCGCGCAAGATCGAGAGGCTGCGCAAGCACTGCTGCGCTGGGCTTTCACCAACCCCGAC 1320  
Db 1261 CCGCGCAAGATCGAGAGGCTGCGCAAGCACTGCTGCGCTGGGCTTTCACCAACCCCGAC 1320

Qy 1321 AAGAAGCACAGAGGAGGCGCCCTTCTGCGCCAT-----CGAGTGCACCCCGACAG 1374  
Db 1321 AAGAAGCACAGAGGAGGCGCCCTTCTGCGCCAT-----CGAGTGCACCCCGACAG 1374

Qy 1375 TGAGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1434  
Db 1375 TGAGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1434

Qy 1435 AGCTGCTGGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCCGGATCAAGTGTGGCGAG 1494  
Db 1435 AGCTGCTGGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCCGGATCAAGTGTGGCGAG 1494

Qy 1495 CTGTGCAAGCTGTGGCGGCGCAGAGGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1554  
Db 1495 CTGTGCAAGCTGTGGCGGCGCAGAGGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1554

Qy 1501 CTGTGCAAGCTGTGGCGGCGCAGAGGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560  
Db 1501 CTGTGCAAGCTGTGGCGGCGCAGAGGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560

Qy 1555 GCGGAGCTGAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1614  
Db 1555 GCGGAGCTGAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1614

Qy 1615 TACGACCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1674  
Db 1615 TACGACCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1674

Qy 1621 TACGACCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680  
Db 1621 TACGACCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680

Qy 1675 TACGAGTCTACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1734

1681 TACCAGATCTACGAGGAGCCCTTCAAGAACCTGAGACCGGACGATGACCCAGATGCGC 1740  
1735 ACCGCCACACACACAGACGCTGAAGCAGCTGACCCGAGGCGGTGCAAGATGCCATGGAG 1794  
1741 ACCGCCACACACACAGACGCTGAAGCAGCTGACCCGAGGCGGTGCAAGATGCCATGGAG 1800  
1795 AGCATCGTGTATCTGGGGCAAGACCCCAAGTTCCCGCTGCCATCCAGAGGAGACCTGG 1854  
1801 AGCATCGTGTATCTGGGGCAAGACCCCAAGTTCCCGCTGCCATCCAGAGGAGACCTGG 1860  
1855 GAGACCTGTGTACACGACTTACTGGCAGGACCACTGGATCCCGGAGTGGAGTTCTGTGAAC 1914  
1861 GAGACCTGTGTACACGACTTACTGGCAGGACCACTGGATCCCGGAGTGGAGTTCTGTGAAC 1920  
1915 ACCCCCCCTCGTGAAGCTGTGTGACAGCTGGAGAGGAGCCCATCATCGGCGCCGAG 1974  
1921 ACCCCCCCTCGTGAAGCTGTGTGACAGCTGGAGAGGAGCCCATCATCGGCGCCGAG 1980  
1975 ACCTTCTAGCTGAGCGGCGCCCAACCGCGAGACCAAGATCGGAGGCGCGGTACGTTG 2034  
1981 ACCTTCTAGCTGAGCGGCGCCCAACCGCGAGACCAAGATCGGAGGCGCGGTACGTTG 2040  
2035 ACCGACCGGGCGCGGAGAGATCGTGAAGCTGTGACCGGAGACCAACCAAGAGACCGAG 2094  
2041 ACCGACCGGGCGCGGAGAGATCGTGAAGCTGTGACCGGAGACCAACCAAGAGACCGAG 2100  
2095 CTGCGAGGCATCCAGCTGCGCTGACGACGACGCGGAGGAGTGAACATCGTGAACGAC 2154  
2101 CTGCGAGGCATCCAGCTGCGCTGACGACGACGCGGAGGAGTGAACATCGTGAACGAC 2160  
2155 AGCCAGTACGCTCGGATCATCCAGGCGCCGAGCCGAGAGGAGGAGGAGTGGTG 2214  
2161 AGCCAGTACGCTCGGATCATCCAGGCGCCGAGAGGAGGAGGAGTGGTG 2220  
2215 AACAGATCATCGAGCAGCTGATCAAGAGGAGGAGTGTACCTGAGCTGGTGGCGCGC 2274  
2221 AACAGATCATCGAGCAGCTGATCAAGAGGAGGAGTGTACCTGAGCTGGTGGCGCGC 2280  
2275 CACAGAGGATCGGCGGCAACGAGCAGATCGAAGCTGTGACCAAGGAGGAGTCCGCAAG 2334  
2281 CACAGAGGATCGGCGGCAACGAGCAGATCGAAGCTGTGACCAAGGAGGAGTCCGCAAG 2340  
2335 GTGCTGTTCTCGACGCGATCGATGGCGGCGATCGTGATCTACCAAGTACATGGAGCCTG 2394  
2341 GTGCTGTTCTCGACGCGATCGATGGCGGCGATCGTGATCTACCAAGTACATGGAGCCTG 2400  
2395 TACGTGGGAGCGGCGGCGCTTAGGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 2451  
2401 TACGTGGGAGCGGCGGCGCTTAGGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 2457

RESULT 9  
US-10-190-305A-38  
; Sequence 38, Application US/10190305A  
; Publication No. US20030198621A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGERDE, Jan  
; APPLICANT: BARNETT, Susan  
; APPLICANT: LIAN, Ying  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
; FILE REFERENCE: 2302-18702 / 18702.002  
; CURRENT APPLICATION NUMBER: US/10/190.305A  
; CURRENT FILING DATE: 2002-07-05  
; NUMBER OF SEQ ID NOS: 93  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 38  
; LENGTH: 2457  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopr.YM.C

US-10-190-305A-38  
Query Match 98.8%; Score 2428.6; DB 14; Length 2457;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 2447; Conservative 0; Mismatches 4; Indels 6; Gaps 1;  
QY 1 GTCCAGCGCACCATGGCGGAGGCCATGAGCCAGCCACAGCCCAACATCTCTGATGCGAG 60  
DB 1 GTCCAGCGCACCATGGCGGAGGCCATGAGCCAGCCACAGCCCAACATCTCTGATGCGAG 60  
QY 61 GCAGCAAACTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCACTGCGGCAAGGAGGC 120  
DB 61 GCAGCAAACTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCACTGCGGCAAGGAGGC 120  
QY 121 CACATCGGCCCAAACTGCGGCGCCCCCGCAAGAGGGCTGTGGAAGTGCAGCAAGGAG 180  
DB 121 CACATCGGCCCAAACTGCGGCGCCCCCGCAAGAGGGCTGTGGAAGTGCAGCAAGGAG 180  
QY 181 GGCACACAGATGAAGAGTGTGACCGAGCGCCAGCCCAAATCTTCTCCGCGAGGACCTGGCC 240  
DB 181 GGCACACAGATGAAGAGTGTGACCGAGCGCCAGCCCAAATCTTCTCCGCGAGGACCTGGCC 240  
QY 241 TTCCCCCAGGSCAAAGGGCCCCGAGTTTCCCCAGGAGCAGAAACCGCGCCCAACAGCCCCACC 300  
DB 241 TTCCCCCAGGSCAAAGGGCCCCGAGTTTCCCCAGGAGCAGAAACCGCGCCCAACAGCCCCACC 300  
QY 301 AGCGCGAGCTGCGAGTGTGCGGGGCGAACCCCGCGAGCGAGCCCGCGCGAGGCCAG 360  
DB 301 AGCGCGAGCTGCGAGTGTGCGGGGCGAACCCCGCGAGCGAGCCCGCGCGAGGCCAG 360  
QY 361 GGCACCTTGAATTTCCCCCAGATCACCTGTGCGAGCGCGCCCCCTGTGAGCATCAAGGTG 420  
DB 361 GGCACCTTGAATTTCCCCCAGATCACCTGTGCGAGCGCGCCCCCTGTGAGCATCAAGGTG 420  
QY 421 GCGCGCGAGTCAAGAGAGGCGCTTCTGAGCACCGGCGCGCGACGACACCTGTGAGGAG 480  
DB 421 GCGCGCGAGTCAAGAGAGGCGCTTCTGAGCACCGGCGCGCGACGACACCTGTGAGGAG 480  
QY 481 ATGAGCTTGGCGGCAAGTGAAGCCCAAGATGATCGGCGGCGATCGCGGCGTTCATCAAG 540  
DB 481 ATGAGCTTGGCGGCAAGTGAAGCCCAAGATGATCGGCGGCGATCGCGGCGTTCATCAAG 540  
QY 541 GTGCGCGAGTACGACGAGTCTGTGAGATCTGCGGCGAGAGGAGCCATCGGCGCGCTG 600  
DB 541 GTGCGCGAGTACGACGAGTCTGTGAGATCTGCGGCGAGAGGAGCCATCGGCGCGCTG 600  
QY 601 CTGATCGGCG 660  
DB 601 CTGATCGGCG 660  
QY 661 ACCCTGAATCTCCCATCGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720  
DB 661 ACCCTGAATCTCCCATCGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720  
QY 721 GACG 780  
DB 721 GACG 780  
QY 781 ATCTCGCGAGGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840  
DB 781 ATCTCGCGAGGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840  
QY 841 AACACCCCGCGTGTTCGCCATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900  
DB 841 AACACCCCGCGTGTTCGCCATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900  
QY 901 TTCCGCGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960  
DB 901 TTCCGCGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960  
QY 961 CCOCGCGCGCTGAAGAGAGAGAGAGCGGTGACCGGTGCTGGAAGGCGGCGAGCGCTTCTC 1020  
DB 961 CCOCGCGCGCTGAAGAGAGAGAGAGCGGTGACCGGTGCTGGAAGGCGGCGAGCGCTTCTC 1020



2095 CTGAGGCCATCCAGCTGGCCCTGCAGGACAGCGGAGCGAGGTGAACATCGTGACCGAC 2154  
2101 CTGAGGCCATCCAGCTGGCCCTGCAGGACAGCGGAGCGAGGTGAACATCGTGACCGAC 2160  
2155 AGCCAGTACGCCCTGGGCATCCTCAGGCCAGCCGACCAAGAGCGAGAGCGAGCTGGTG 2214  
2161 AGCCAGTACGCCCTGGGCATCCTCAGGCCAGCCGACCAAGAGCGAGAGCGAGCTGGTG 2220  
2215 AACAGATCATCGAGCAGCTGATCAAGAGAGAGGTGATCCTGAGCTGGTGCCCGCC 2274  
2221 AACAGATCATCGAGCAGCTGATCAAGAGAGAGGTGATCCTGAGCTGGTGCCCGCC 2280  
2275 CACAAGGCGATCGCGGCGCAAGCAGATCGACAAGCTGGTGAGCAAGGGGATCCGCAAG 2334  
2281 CACAAGGCGATCGCGGCGCAAGCAGATCGACAAGCTGGTGAGCAAGGGGATCCGCAAG 2340  
2335 GTGCTGTTCTTGGAGCGGCATCGATGGCGGCATCGTGTATCTACAGTACATGGAGCACTG 2394  
2341 GTGCTGTTCTTGGAGCGGCATCGATGGCGGCATCGTGTATCTACAGTACATGGAGCACTG 2400  
2395 TACGTGGGACGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451  
2401 TACGTGGGACGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

RESULT 10  
US-10-190-435-13  
; Sequence 13, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEHDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Esterlita J.  
; TITLE OF INVENTION: POLYPEPTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: P18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 3531  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GagPolmut\_C  
US-10-190-435-13

Query Match 97.9%; Score 2404.4; DB 14; Length 3531;  
Best Local Similarity 99.1%; Pred. No. 0;  
Matches 2417; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

14 TGCCCGAGCGCATGAGCCAGGCGCAACAGCGCCCAATCTGATGAGCGGCAACTTCA 73  
1088 TGCCCGAGCGCATGAGCCAGGCGCAACAGCGCCCAATCTGATGAGCGGCAACTTCA 1147  
74 AGGGCCCCAAGCGCATCATCAAGTCTTCAACTGGCGCAAGAGGGCCACATCGCCGCA 133  
1148 AGGGCCCCAAGCGCATCATCAAGTCTTCAACTGGCGCAAGAGGGCCACATCGCCGCA 1207  
134 ACTGCGCGCGCCCCCGCAAGAGGGCTGTGAAAGTGGCGCAAGAGGGCCACAGATGA 193  
1208 ACTGCGCGCGCCCCCGCAAGAGGGCTGTGAAAGTGGCGCAAGAGGGCCACAGATGA 1267  
194 AGGACTGACCGAGCGCCAGGCGCAACTCTTCCGAGGAGCCTGGCTTCCCGCGGCA 253  
1268 AGGACTGACCGAGCGCCAGGCGCAACTCTTCCGAGGAGCCTGGCTTCCCGCGGCA 1327  
254 AGGCGCGCGAGTTTCCCGAGCGAGCAGACCGCGCCCAAGAGGGCCACAGCGGAGCTGC 313  
1328 AGGCGCGCGAGTTTCCCGAGCGAGCAGACCGCGCCCAAGAGGGCCACAGCGGAGCTGC 1387

1021 AGCGTCCCTTGGACGAGGACTTCCGCAAGTACACCGCTTCCACATCCCGAGCATCAAC 1080  
1021 AGCGTCCCTTGGACGAGGACTTCCGCAAGTACACCGCTTCCACATCCCGAGCATCAAC 1080  
1081 AACGAGACCCCGCGCATCCGCTACCAAGTACAACGCTGCTGCCCGAGGGCTGGAAGGCGAC 1140  
1081 AACGAGACCCCGCGCATCCGCTACCAAGTACAACGCTGCTGCCCGAGGGCTGGAAGGCGAC 1140  
1141 CCAGCATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCGCGCAACCCC 1200  
1141 CCAGCATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCGCGCAACCCC 1200  
1201 GAGATCGTGTATACCAAGGCGCCCTTCTGCTGCCCAT-----CGAGCTGCAACCCCGCAAG 1374  
1321 AAGNAGCACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380  
1321 AAGNAGCACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380  
1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1434  
1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440  
1435 AAGCTGCTGGCAGGCTGAAGTGGCGCAGCAGATCTACCCGCGCATCAAGGTGGCGGAG 1494  
1441 AAGCTGCTGGCAGGCTGAAGTGGCGCAGCAGATCTACCCGCGCATCAAGGTGGCGGAG 1500  
1495 CTGTGAAGTGTCTGCGCGCGCCCAAGCGCTGACCGACATCGTGGCCCTGACCGAGGAG 1554  
1501 CTGTGAAGTGTCTGCGCGCGCCCAAGCGCTGACCGACATCGTGGCCCTGACCGAGGAG 1560  
1555 GCCAGCTGGAGCTGGCGGAGACCGCGAGATCTGCGCGAGCCCGTGCACGGCGGTGAC 1614  
1561 GCCAGCTGGAGCTGGCGGAGACCGCGAGATCTGCGCGAGCCCGTGCACGGCGGTGAC 1620  
1615 TACGACCCGACAGAGGAGCTGTGCGCGAGATCCAGAGGAGGAGGAGGAGGAGGAGGAG 1674  
1621 TACGACCCGACAGAGGAGCTGTGCGCGAGATCCAGAGGAGGAGGAGGAGGAGGAGGAG 1680  
1675 TACGAGATCTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1734  
1681 TACGAGATCTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740  
1735 ACCGCGCACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1794  
1741 ACCGCGCACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800  
1795 AGCATCGTGTATCTGGGCGAAGACCCCGCAAGTTCCGCTGCGCCATCCAGAGAGGAGCTGG 1854  
1801 AGCATCGTGTATCTGGGCGAAGACCCCGCAAGTTCCGCTGCGCCATCCAGAGAGGAGCTGG 1860  
1855 GAGACTGGTGGACCGCATCTGGAGGCGCAGCTGGATCCCGAGTGGAGTTCGTGAC 1914  
1861 GAGACTGGTGGACCGCATCTGGAGGCGCAGCTGGATCCCGAGTGGAGTTCGTGAC 1920  
1915 ACCCGCGCGCTGCTGAGTGTGTACAGCTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1974  
1921 ACCCGCGCGCTGCTGAGTGTGTACAGCTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980  
1975 ACCTTTACTAGTGGAGCGGCGCGCCCAAGCGGAGACCAAGATCGGCAAGCGCGGCTACGTG 2034  
1981 ACCTTTACTAGTGGAGCGGCGCGCCCAAGCGGAGACCAAGATCGGCAAGCGCGGCTACGTG 2040  
2035 ACCGACCGGCGCGCGAGAGTCTGTGAGCTGTGACCGGAGACCAAGGAGGAGGAGGAGGAG 2094  
2041 ACCGACCGGCGCGCGAGAGTCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAG 2100

QY 314 AGGTGCGCGGACAAACCCCGCAGCGAGCGCGCGAGCGCCAGCGCACCCCTGAACT 373  
Db 1388 AGGTGCGCGGACAAACCCCGCAGCGAGCGCGCGAGCGCCAGCGCACCCCTGAACT 1447  
QY 374 TCCCCCAAGTCAACCTGTGGAGCGCGCCCTGTGTAGCATCAAGTGGCGCGCAGATCA 433  
Db 1448 TCCCCCAAGTCAACCTGTGGAGCGCGCCCTGTGTAGCATCAAGTGGCGCGCAGATCA 1507  
QY 434 AGGAGCCCTGTGTGACACCGCGCGCGAGCACCGCTGTGTGAGGAGATGAGCTTCCCG 493  
Db 1508 AGGAGCCCTGTGTGACACCGCGCGCGAGCACCGCTGTGTGAGGAGATGAGCTTCCCG 1567  
QY 494 GCAAGTGGAGCCCAAGATGATCGCGGGCATCGCGGGCTTCAATCAAGTGGCGCGCAGTACG 553  
Db 1568 GCAAGTGGAGCCCAAGATGATCGCGGGCATCGCGGGCTTCAATCAAGTGGCGCGCAGTACG 1627  
QY 554 ACCAGATCTGTGAGATCTCGGCAAGAAAGCCCATCGGCACCGCTGTGTGATCGGCCCA 613  
Db 1628 ACCAGATCTGTGAGATCTCGGCAAGAAAGCCCATCGGCACCGCTGTGTGATCGGCCCA 1687  
QY 614 CCGCGTGAACATCATCGCGCGCAACATGCTGACCGAGCTGGGCTGCAACCTGAACTTCC 673  
Db 1688 CCGCGTGAACATCATCGCGCGCAACATGCTGACCGAGCTGGGCTGCAACCTGAACTTCC 1747  
QY 674 CCATCAGCCCCATCGAGACCGCTGCAAGCTGAAGCCCGCATGAGCGGCCCAAGG 733  
Db 1748 CCATCAGCCCCATCGAGACCGCTGCAAGCTGAAGCGGCCCATGAGCGGCCCAAGG 1807  
QY 734 TGAAGCAGTGGCCCTGTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793  
Db 1808 TGAAGCAGTGGCCCTGTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 1867  
QY 794 TGGAGAGGAGGCGAGATCAACAGATCGCGCGCGAGAACCCCTACACACCCCGCTGT 853  
Db 1868 TGGAGAGGAGGCGAGATCAACAGATCGCGCGCGAGAACCCCTACACACCCCGCTGT 1927  
QY 854 TCGCCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGA 913  
Db 1928 TCGCCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGA 1987  
QY 914 ACAGCGCACCCAGAGCTTCTGGAGGTGAGCTGGGCTACCGATCCCGACCCCGCGCTGA 973  
Db 1988 ACAGCGCACCCAGAGCTTCTGGAGGTGAGCTGGGCTACCGATCCCGACCCCGCGCTGA 2047  
QY 974 AGAAGAAAGAGCGCTGACCGCTGCTGGACGTGGCGAGCGCTTCAAGCGTGGCCCTGG 1033  
Db 2048 AGAAGAAAGAGCGCTGACCGCTGCTGGACGTGGCGAGCGCTTCAAGCGTGGCCCTGG 2107  
QY 1034 ACAGAGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAGAGACCGCG 1093  
Db 2108 ACAGAGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAGAGACCGCG 2167  
QY 1094 GCATCCGCTACCAAGTACACCGCTTCCCGCAGGGCTGGAGGGCAGCCCGACATCTTCC 1153  
Db 2168 GCATCCGCTACCAAGTACACCGCTTCCCGCAGGGCTGGAGGGCAGCCCGACATCTTCC 2227  
QY 1154 AGAGCAGCATGACCAAGATCTTGGAGCGCTTCCCGCGCCCGCAACCCCGAGATCGTATCT 1213  
Db 2228 AGAGCAGCATGACCAAGATCTTGGAGCGCTTCCCGCGCCCGCAACCCCGAGATCGTATCT 2287  
QY 1214 ACCAGCGCCCTGTGTAGTGGGACGCACTTGGAGATCGCGCAGCACCCGCGCAAGATCG 1273  
Db 2288 ACCAGCGCCCTGTGTAGTGGGACGCACTTGGAGATCGCGCAGCACCCGCGCAAGATCG 2347  
QY 1274 AGAGCTGCGCAAGCAGCTGCTGCGTGGGGCTTCAACACCCCGCAAGAGACCCAGA 1333  
Db 2348 AGAGCTGCGCAAGCAGCTGCTGCGTGGGGCTTCAACACCCCGCAAGAGACCCAGA 2407  
QY 1334 AGAGCGCCCTTCCCTGCGCCATCAGCTGCACCCCGCAAGTGGAGCGCTGAGCGCCATCG 1393  
Db 2408 AGAGCGCCCTTCCCTGCGCCATCAGCTGCACCCCGCAAGTGGAGCGCTGAGCGCCATCG 2467

QY 1394 AGTGCCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAGCTGTGGCGAGCTGA 1453  
Db 2468 AGTGCCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAGCTGTGGCGAGCTGA 2527  
QY 1454 ACTGGGCCAGCCAGATCTACCCCGCATCAAGGTGGCGAGCTGTGCAAGCTGTGCGCG 1513  
Db 2528 ACTGGGCCAGCCAGATCTACCCCGCATCAAGGTGGCGAGCTGTGCAAGCTGTGCGCG 2587  
QY 1514 GCGCCAAAGCGCTGACCGGACATCGTCCCTGACCGAGAGGCCAGCTGTGAGCTGGCG 1573  
Db 2588 GCGCCAAAGCGCTGACCGGACATCGTCCCTGACCGAGAGGCCAGCTGTGAGCTGGCG 2647  
QY 1574 AGAACCGCGAGATCTTGGCGGAGCCGCTGCAAGGGGTGTACTACGACCCCGACAGGACC 1633  
Db 2648 AGAACCGCGAGATCTTGGCGGAGCCGCTGCAAGGGGTGTACTACGACCCCGACAGGACC 2707  
QY 1634 TGTGCGCGAGATCCAGAAAGAGGGGCCAGACGACGAGTGAACCTACCGATCTACGAGAGC 1693  
Db 2708 TGTGCGCGAGATCCAGAAAGAGGGGCCAGACGAGTGAACCTACCGATCTACGAGAGC 2767  
QY 1694 CCTTCAGAACCTGAAGACCGGCAAGTACGCAAGATGCGCACCGGCCACACCAACGAG 1753  
Db 2768 CCTTCAGAACCTGAAGACCGGCAAGTACGCAAGATGCGCACCGGCCACACCAACGAG 2827  
QY 1754 TGAAGCAGCTGACCGAGGGCGGTGCAAGAGATCGCCATGAGAGCATCTGTGATCTGGGGCA 1813  
Db 2828 TGAAGCAGCTGACCGAGGGCGGTGCAAGAGATCGCCATGAGAGCATCTGTGATCTGGGGCA 2887  
QY 1814 AGACCCCGAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGTGTGACCGACT 1873  
Db 2888 AGACCCCGAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGTGTGACCGACT 2947  
QY 1874 ACTGGCAGGGCACCTGTGATCCCGAGTGGAGTTCGTGAACACCCCGCTGTGTGAAGC 1933  
Db 2948 ACTGGCAGGGCACCTGTGATCCCGAGTGGAGTTCGTGAACACCCCGCTGTGTGAAGC 3007  
QY 1934 TGTGTTACAGCTTGGAGAGGAGCCCATCATCGGCGCGAGACCTGTGAGGCGATCCAGCTGG 1993  
Db 3008 TGTGTTACAGCTTGGAGAGGAGCCCATCATCGGCGCGAGACCTGTGAGGCGATCCAGCTGG 3067  
QY 1994 CCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACCCGCGCGCGCAGA 2053  
Db 3068 CCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACCCGCGCGCGCAGA 3127  
QY 2054 AGATCGTGAAGCTTGAACGAGACCAACCAAGAGAGCCGAGCTGTGAGGCGATCCAGCTGG 2113  
Db 3128 AGATCGTGAAGCTTGAACGAGACCAACCAAGAGAGCCGAGCTGTGAGGCGATCCAGCTGG 3187  
QY 2114 CCCTGAGGACAGCGCGCAGCGAGTGAACATCTGTGACCGACAGCCAGCTACGCCCTGGGCA 2173  
Db 3188 CCCTGAGGACAGCGCGCAGCGAGTGAACATCTGTGACCGACAGCCAGCTACGCCCTGGGCA 3247  
QY 2174 TCATCCAGGCCCGCGCAAGAGCGAGAGCGAGCTGTGAAACAGATCATCGAGCAGC 2233  
Db 3248 TCATCCAGGCCCGCGCAAGAGCGAGAGCGAGCTGTGAAACAGATCATCGAGCAGC 3307  
QY 2234 TGATCAAGAGGAGAGAGTGTACTGTAGCTGGGTGCCCGCCCAAGAGGCGATCGGCGGCA 2293  
Db 3308 TGATCAAGAGGAGAGAGTGTACTGTAGCTGGGTGCCCGCCCAAGAGGCGATCGGCGGCA 3367  
QY 2294 ACAGCAGATCGACAGCTGTGAGCAAGGGCATTCGCAAGGTGTGTTCTTCTGGAGCGCA 2353  
Db 3368 ACAGCAGATCGACAGCTGTGAGCAAGGGCATTCGCAAGGTGTGTTCTTCTGGAGCGCA 3427  
QY 2354 TCGATGGCGGCGATCGTATCTACCATGACGACCTGTACGTGGGCGAGCGCGGCGC 2413  
Db 3428 TCGATGGCGGCGATCGTATCTACCATGACGACCTGTACGTGGGCGAGCGCGGCGC 3487  
QY 2414 CTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2451  
Db 3488 CTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 3525





Db 1741 ACCGCCCAACACACAGCTGAGCAGCTGACCGGCGGTGACAGATGCCATGGAG 1800  
Qy 1795 AGCATGTGATCTGGGGGAAGACCCCAAGTTCCGCTGCCATCCAGAGGAGACCTGG 1854  
Db 1801 AGCATGTGATCTGGGGGAAGACCCCAAGTTCCGCTGCCATCCAGAGGAGACCTGG 1860  
Qy 1855 GAGACCTGTGAGACCACTATCTGGCAGGCACTGTGATCCCGAGTGGGAGTTCGTGAAC 1914  
Db 1861 GAGACCTGTGAGACCACTATCTGGCAGGCACTGTGATCCCGAGTGGGAGTTCGTGAAC 1920  
Qy 1915 ACCCCCCCTGTGAGCTGTGTTACAGCTGGAGAGGACCCATCATCCGCGCCGAG 1974  
Db 1921 ACCCCCCCTGTGAGCTGTGTTACAGCTGGAGAGGACCCATCATCCGCGCCGAG 1980  
Qy 1975 ACCTTTCTAGTGAACCGGCGCCCAACCGCGAGACCAAGTGGGCAAGGCCGCTACGTG 2034  
Db 1981 ACCTTTCTAGTGAACCGGCGCCCAACCGCGAGACCAAGTGGGCAAGGCCGCTACGTG 2040  
Qy 2035 ACCGACGGGGCGGCGAGAGATCGTGAGCTGACCGAGACCAACAGAGACCGGAG 2094  
Db 2041 ACCGACGGGGCGGCGAGAGATCGTGAGCTGACCGAGACCAACAGAGACCGGAG 2100  
Qy 2095 CTCAGGCCATTCAGCTGGCCCTGACAGCAGCGGAGCGAGGTGAACATCGTGACCGAC 2154  
Db 2101 CTCAGGCCATTCAGCTGGCCCTGACAGCAGCGGAGCGAGGTGAACATCGTGACCGAC 2160  
Qy 2155 AGCAGTACCCCTGGGCATCTCCAGGCCCGCCAGCGCAAGAGCGAGCGAGCTGGTG 2214  
Db 2161 AGCAGTACCCCTGGGCATCTCCAGGCCCGCCAGCGCAAGAGCGAGCGAGCTGGTG 2220  
Qy 2215 AACAGATCATCAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCCCGCC 2274  
Db 2221 AACAGATCATCAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCCCGCC 2280  
Qy 2275 CACAGGGCATCGGCGCACGAGCAGATCGACAAGCTGTGAGCAGGGCATCCGCAAG 2334  
Db 2281 CACAGGGCATCGGCGCACGAGCAGATCGACAAGCTGTGAGCAGGGCATCCGCAAG 2340  
Qy 2335 GTGCTGTCTGAGCGGCATCGATGGCGGCATCGTGATCTTACAGTACATGAGCAGCTG 2394  
Db 2341 GTGCTGTCTGAGCGGCATCGATGGCGGCATCGTGATCTTACAGTACATGAGCAGCTG 2400  
Qy 2395 TACGTGGGAGCGCGCCCTAGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 2451  
Db 2401 TACGTGGGAGCGCGCCCTAGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 2457

RESULT 12

US-10-190-305A-39  
; Sequence 39, Application US/10190305A  
; Publication No. US20030198621A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEDEDE, Jan  
; APPLICANT: BARNETT, Susan  
; APPLICANT: LIAN, Ying  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
; TYPE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: 2302-18702 / 18702.002  
; CURRENT APPLICATION NUMBER: US/10/190,305A  
; CURRENT FILING DATE: 2002-07-05  
; NUMBER OF SEQ ID NOS: 93  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 39  
; LENGTH: 2457  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt\_C  
US-10-190-305A-39

Query Match 97.8%; Score 2403.4; DB 14; Length 2457;  
Best Local Similarity 99.3%; Pred. No. 0;

Matches 2439; Conservative 0; Mismatches 6; Indels 12; Gaps 2;  
Qy 7 GCCACCATGCGCCGAGGCGCATGAGCCAGGCGCAACAGCGCCCAACATCTGTGATGAGCGCGCAGC 66  
Db 1 GCCACCATGCGCCGAGGCGCATGAGCCAGGCGCAACAGCGCCCAACATCTGTGATGAGCGCGCAGC 60  
Qy 67 AACTTCAAGGGGCCCCCAAGGCGCATCATCAAGTGTCTTCAACTGGGGCAAGGAGGGCCACATC 126  
Db 61 AACTTCAAGGGGCCCCCAAGGCGCATCATCAAGTGTCTTCAACTGGGGCAAGGAGGGCCACATC 120  
Qy 127 GCCCGCAACTGCGCGCCCGCCCGCAAGAGGGGTGTGTGAAGTGTGCGGCAAGGAGGGCCAC 186  
Db 121 GCCCGCAACTGCGCGCCCGCCCGCAAGAGGGGTGTGTGAAGTGTGCGGCAAGGAGGGCCAC 180  
Qy 187 CAGATGAAGGACTGCAACCGAGCGCCAGGCGCAACTTCTTCGCGAGAGACCTTGGCCTTCCCGC 246  
Db 181 CAGATGAAGGACTGCAACCGAGCGCCAGGCGCAACTTCTTCGCGAGAGACCTTGGCCTTCCCGC 240  
Qy 247 CAGGGCAAGGCCCGCGAGTTCCCGAGCGAGCAGAAACCGCGCCCAACAGCGCCCAACAGCGCGC 306  
Db 241 CAGGGCAAGGCCCGCGAGTTCCCGAGCGAGCAGAAACCGCGCCCAACAGCGCCCAACAGCGCGC 300  
Qy 307 GAGCTGCAAGTGTGGCGGCGAGCAACCCCGCAGCGAGCGCGCGCCCGAGCGCCAGGGGCAAC 366  
Db 301 GAGCTGCAAGTGTGGCGGCGAGCAACCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 360  
Qy 367 CTGAACCTTCCCGCAGATCAACCTTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 426  
Db 361 CTGAACCTTCCCGCAGATCAACCTTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 420  
Qy 427 CAGATCAAGGAGGCCCTGTCTGGACACCGGGCGCGAGCACCGCTGTGGAGGAGATGAGC 486  
Db 421 CAGATCAAGGAGGCCCTGTCTGGACACCGGGCGCGAGCACCGCTGTGGAGGAGATGAGC 480  
Qy 487 CTGCGCGCAAGTGTGAAGCGCCCAAGATGATCGGGCGCATTCGGCGGTTCATCAAGAGTGGCG 546  
Db 481 CTGCGCGCAAGTGTGAAGCGCCCAAGATGATCGGGCGCATTCGGCGGTTCATCAAGAGTGGCG 540  
Qy 547 CAGTACGACCAAGATCTGTGATGAGATCTGCGGCAAGAGGCGCATCGCACCGTGTGATC 606  
Db 541 CAGTACGACCAAGATCTGTGATGAGATCTGCGGCAAGAGGCGCATCGCACCGTGTGATC 600  
Qy 607 GCGCCCGCCCGCGTGAACATCATCGCGCGCAACATGTGTACCCAGCTGGGCTGCGACCGCTG 666  
Db 601 GCGCCCGCCCGCGTGAACATCATCGCGCGCAACATGTGTGTACCCAGCTGGGCTGCGACCGCTG 660  
Qy 667 AACTTCCCGCATCGCCCGCATCGAGACCGTGTGAGCTGAGCGCGCGCGCGCGCGCGCGCGC 726  
Db 661 AACTTCCCGCATCGCCCGCATCGAGACCGTGTGAGCTGAGCGCGCGCGCGCGCGCGCGCGC 720  
Qy 727 CCCAAGGTGAAGCAGTGTGCGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCGCATCTGC 786  
Db 721 CCCAAGGTGAAGCAGTGTGCGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCGCATCTGC 780  
Qy 787 GAGGAGATGGAAGAGGAGGCGCAAGATCAACAGATCGGCGCCCGGAGACCCCTACCAACACC 846  
Db 781 GAGGAGATGGAAGAGGAGGCGCAAGATCAACAGATCGGCGCCCGGAGACCCCTACCAACACC 840  
Qy 847 CCGGTGTTCGCCATCAAGAGAGAGCAGACCAAGTGGCGCGCAAGCTGGTGAATCTCCGC 906  
Db 841 CCGGTGTTCGCCATCAAGAGAGAGCAGACCAAGTGGCGCGCAAGCTGGTGAATCTCCGC 900  
Qy 907 GAGCTGAACAAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCGCACCCCGCC 966  
Db 901 GAGCTGAACAAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCGCACCCCGCC 960  
Qy 967 GCGCTGAAGAGAGAGAGGCGTGTGCGCTGTGAGCGCGCGCGCGCTTCTCAGCGTG 1026  
Db 961 GCGCTGAAGAGAGAGAGGCGTGTGCGCTGTGAGCGCGCGCGCGCTTCTCAGCGTG 1020  
Qy 1027 CCGCTTGGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAACAAACGAG 1086  
Db 1021 CCGCTTGGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAACAAACGAG 1080

QY 1087 ACCCCGGGATCCGCTACCAAGTCTGCTCCCGCAGGGTGGAGGGCGACGCCAGC 1146  
Db 1081 ACCCCGGGATCCGCTACCAAGTCTGCTCCCGCAGGGTGGAGGGCGACGCCAGC 1140  
QY 1147 ATCTTCAGAGAGCATGACCAAGATCTCGAGCCCTTCGGCGCCGCAACCCGAGATC 1206  
Db 1141 ATCTTCAGAGAGCATGACCAAGATCTCGAGCCCTTCGGCGCCGCAACCCGAGATC 1200  
QY 1207 GTGATCTACCA-----GGCCCGCCTGTAGTGGGAGAGCACTCGAGATCGGCCAGCAC 1260  
Db 1201 GTGATCTACCAAGTACGAGGAGCTGTAGTGGGAGAGCACTCGAGATCGGCCAGCAC 1260  
QY 1261 CGGCCAAGATCGAGAGAGTGGCAAGCACTGTGTGGGAGAGCTTACCAACCCCGAG 1320  
Db 1261 CGGCCAAGATCGAGAGAGTGGCAAGCACTGTGTGGGAGAGCTTACCAACCCCGAG 1320  
QY 1321 AAGAAGCACAGAGAGAGCCCGCTTCTGTCCCAT-----CGAGGTGCAACCCCGACAAG 1374  
Db 1321 AAGAAGCACAGAGAGAGCCCGCTTCTGTGGATGGCTACGAGCTGCAACCCCGACAAG 1380  
QY 1375 TGGACCGTGCAGGCCATCGAGTCTCCCGAGAGAGAGAGCTGGACCGTGAACGACATCCAG 1434  
Db 1381 TGGACCGTGCAGGCCATCGAGTCTCCCGAGAGAGAGAGCTGGACCGTGAACGACATCCAG 1440  
QY 1435 AAGCTGTGGCAAGCTGAAGTGGGCGCAGCAGATCTACCCCGCATCAAGGTGGCCAG 1494  
Db 1441 AAGCTGTGGCAAGCTGAAGTGGGCGCAGCAGATCTACCCCGCATCAAGGTGGCCAG 1500  
QY 1495 CTGTGAAGCTGTGCGCGCGCCCAAGCGCCTGACCAATCGTGTGCCCTTGACCGAGGAG 1554  
Db 1501 CTGTGAAGCTGTGCGCGCGCCCAAGCGCCTGACCAATCGTGTGCCCTTGACCGAGGAG 1560  
QY 1555 GCCAGCTGGAGTGGCGGAGACCGGAGATCTGTGGCGAGCCCGGCGACGGGTGTAC 1614  
Db 1561 GCCAGCTGGAGTGGCGGAGACCGGAGATCTGTGGCGAGCCCGGCGACGGGTGTAC 1620  
QY 1615 TACGACCCCGACAGAGAGCTGTGTGGCGAGATCCAGAGAGAGGCGCCACGACGTGAGCC 1674  
Db 1621 TACGACCCCGACAGAGAGCTGTGTGGCGAGATCCAGAGAGAGGCGCCACGACGTGAGCC 1680  
QY 1675 TACGAGTCTACGAGAGCCCTTCAAGAACCTGAGACCGGCAAGTACGCGAGATGCGC 1734  
Db 1681 TACGAGTCTACGAGAGCCCTTCAAGAACCTGAGACCGGCAAGTACGCGAGATGCGC 1740  
QY 1735 ACCGCCACACAGAGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGAGTCCGCATGGAG 1794  
Db 1741 ACCGCCACACAGAGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGAGTCCGCATGGAG 1800  
QY 1795 AGCATCGTGTATCTGGGGCAAGACCCCAAGTTCCGCTTCCCATCCAGAGGAGACCTGG 1854  
Db 1801 AGCATCGTGTATCTGGGGCAAGACCCCAAGTTCCGCTTCCCATCCAGAGGAGACCTGG 1860  
QY 1855 GAGACCTGGTGGACCGACTACTGGACGCCACCTGGATCCCGAGTGGGAGTTGTTGAAC 1914  
Db 1861 GAGACCTGGTGGACCGACTACTGGACGCCACCTGGATCCCGAGTGGGAGTTGTTGAAC 1920  
QY 1915 ACCCCCCCTGTGTAGCTGTGTACAGCTGGAGAGAGGCCATCATCGCGCCGAG 1974  
Db 1921 ACCCCCCCTGTGTAGCTGTGTGTACAGCTGGAGAGAGGCCATCATCGCGCCGAG 1980  
QY 1975 ACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTG 2034  
Db 1981 ACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTG 2040  
QY 2035 ACCGACCGGGCGCGCAGAGATCGTGTAGCTGACCGAGAGACCAACAGAGAGAGCGAG 2094  
Db 2041 ACCGACCGGGCGCGCAGAGATCGTGTAGCTGACCGAGAGACCAACAGAGAGAGCGAG 2100  
QY 2095 CTGACGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGAGGTGAACATCGTGACCGAC 2154  
Db 2101 CTGACGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGAGGTGAACATCGTGACCGAC 2160

QY 2155 AGCCAGTACGCCCTGGGATCATCCAGGCCAGCCGACAGAGCGAGAGCGAGCTGGTG 2214  
Db 2161 AGCCAGTACGCCCTGGGATCATCCAGGCCAGCCGACAGAGCGAGAGCGAGCTGGTG 2220  
QY 2215 AACAGATCATCGAGAGCTGTATCAAGAGAGAGAGTGTACTGTAGCTGGTGGTCCGCC 2274  
Db 2221 AACAGATCATCGAGAGCTGTATCAAGAGAGAGAGTGTACTGTAGCTGGTGGTCCGCC 2280  
QY 2275 CACAGGGCATCGCGCGGACAGCAGATCGAAGAGTGTGTAGCAAGGGCATCCGCAAG 2334  
Db 2281 CACAGGGCATCGCGCGGACAGCAGATCGAAGAGTGTGTAGCAAGGGCATCCGCAAG 2340  
QY 2335 GTGCTGTTCCTGGACGGCATCGATGGGGCATCTGTATCTACCAAGTATCATGACGACCTG 2394  
Db 2341 GTGCTGTTCCTGGACGGCATCGATGGGGCATCTGTATCTACCAAGTATCATGACGACCTG 2400  
QY 2395 TACGTGGCAGCGCGGCCCTAGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 2451  
Db 2401 TACGTGGCAGCGCGGCCCTAGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 2457

RESULT 13  
US-10-190-435-14  
; Sequence 14, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Esclrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: PP18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 3537  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GagPolmutAtt\_C  
US-10-190-435-14

Query Match 97.8%; Score 2402.8; DB 14; Length 3537;  
Best Local Similarity 99.1%; Pred. No. 0;  
Matches 2416; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 14 TGGCCGAGCCATGAGCCAGGCCACAGCGCCAAATCTCTGTATGCGAGCGCAACTTCA 73  
Db 1094 TGGCCGAGCCATGAGCCAGGCCACAGCGCGTATGATGAGAGAGCAACTTTAAAA 1153  
QY 74 AGGCCCCCAAGCGCATCATCAAGTCTTCACTGCGGCAAGAGAGGCCACATCGCCGCA 133  
Db 1154 AGGCCCCCAAGCGCATCATCAAGTCTTCACTGCGGCAAGAGAGGCCACATCGCCGCA 1213  
QY 134 ACTGCGCGCCCCCGCCAGAGAGGCTGTGTGAAGTGGGGCAAGAGAGGCCCAACAGATGA 193  
Db 1214 ACTGCGCGCCCCCGCCAGAGAGGCTGTGTGAAGTGGGGCAAGAGAGGCCCAACAGATGA 1273  
QY 194 AGGACTGCAACGAGCGCCAGGCGCAACTTCTTCGGGAGGACCTTGCCCTTCCCGAGGCA 253  
Db 1274 AGGACTGCAACGAGCGCCAGGCGCAACTTCTTCGGGAGGACCTTGCCCTTCCCGAGGCA 1333  
QY 254 AGSCCGCGAGTTCCTCCAGCAGAGCAACCCGCGCAACAGCCGCCACAGCCCGAGCTGC 313  
Db 1334 AGSCCGCGAGTTCCTCCAGCAGAGCAACCCGCGCAACAGCCGCCACAGCCCGAGCTGC 1393  
QY 314 AGTGTGCGCGCAACAAACCCCGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 373  
Db 1394 AGTGTGCGCGCAACAAACCCCGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1453

QY	374	TCCCCCAGATCACCCCTGTGGAGCGCCCTTGGTGGAGATCAAGGTGGCGCGCCAGATCA	433	1454	ACTGGGCCAGCCAGATCTTACCCCGGCATCAAGGTGGCGCCAGCTGTGCAGAGCTGTGCGCG	1513
Db	1454	TCCCCCAGATCACCCCTGTGGAGCGCCCTTGGTGGAGATCAAGGTGGCGCGCCAGATCA	1513	2534	ACTGGGCCAGCCAGATCTTACCCCGGCATCAAGGTGGCGCCAGCTGTGCAGAGCTGTGCGCG	2593
QY	434	AGGAGCCCTGTGTGGACACCGCGCGCGACACACCGTGTCTGGAGGAGATGAGCTTGCCCG	493	1514	CGCCCAAGGCCCTGACCGACATCGTCCCTGACCGAGAGGCCGAGCTGGAGCTGGCCCG	1573
Db	1514	AGGAGCCCTGTGTGGACTCCGCGCGCGACACACCGTGTCTGGAGGAGATGAGCTTGCCCG	1573	2594	CGCCCAAGGCCCTGACCGACATCGTCCCTGACCGAGAGGCCGAGCTGGAGCTGGCCCG	2653
QY	494	GCAAGTGGAAAGCCCAAGATGATCGCGCGATCGCGCGCTTCAATCAAGGTGGCGCGAGTAG	553	1574	AGAACCCGAGATCTCTGGCGAGCCCGTGTACCGGCTGTACTACGACCCCGCAGCAGGACC	1633
Db	1574	GCAAGTGGAAAGCCCAAGATGATCGCGCGATCGCGCGCTTCAATCAAGGTGGCGCGAGTAG	1633	2654	AGAACCCGAGATCTCTGGCGAGCCCGTGTACCGGCTGTACTACGACCCCGCAGCAGGACC	2713
QY	554	ACCAGATCTCTGATCGAGATCTGGCGCAAGAAGGCCATCGGCACCCGTGTGATCGGCCCA	613	1634	TGTTGGCCGAGATTCAGAACCGGCGCCACGACAGTGGACCTACCAAGATCTACCAAGAGC	1693
Db	1634	ACCAGATCTCTGATCGAGATCTGGCGCAAGAAGGCCATCGGCACCCGTGTGATCGGCCCA	1693	2714	TGTTGGCCGAGATTCAGAACCGGCGCCACGACAGTGGACCTACCAAGATCTACCAAGAGC	2773
QY	614	CCCCGTGAAATCATCTCGCGCGCAACATGCTGACCCAGCTGGCTGGCTGACCTGAACTTCC	673	1694	CTTTCAAGAACTCTGAAGACCCGCAAGTACGCCAAGATGCGCACCGGCCACACCAACGACG	1753
Db	1694	CCCCGTGAAATCATCTCGCGCGCAACATGCTGACCCAGCTGGCTGGCTGACCTGAACTTCC	1753	2774	CTTTCAAGAACTCTGAAGACCCGCAAGTACGCCAAGATGCGCACCGGCCACACCAACGACG	2833
QY	674	CCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATGAGCGGCCCAAGG	733	1754	TGAAGAGCTGACCGAGGCCGTGACAGAGATCGCCATCGGAGAGATCTGTGATCTGGGGCA	1813
Db	1754	CCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATGAGCGGCCCAAGG	1813	2834	TGAAGAGCTGACCGAGGCCGTGACAGAGATCGCCATCGGAGAGATCTGTGATCTGGGGCA	2893
QY	734	TGAAGAGCTGAGCCCTGACCGAGAGAAATCAAGGCCCTGACCGCCATCTGGAGGAGA	793	1814	AGACCCCAAGTTCGGCTGCCATCCAGAAAGGAGACCTGGGAGACCTGTGTGACCGACT	1873
Db	1814	TGAAGAGCTGAGCCCTGACCGAGAGAAATCAAGGCCCTGACCGCCATCTGGAGGAGA	1873	2894	AGACCCCAAGTTCGGCTGCCATCCAGAAAGGAGACCTGGGAGACCTGTGTGACCGACT	2953
QY	794	TGAGAGAGGAGGCAAGATCAACAGATCGGCCCGGAGAACCCCTACAAACACCCCGTGT	853	1874	ACTGGCAGGCCACTTGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGTGTGAAGC	1933
Db	1874	TGAGAGAGGAGGCAAGATCAACAGATCGGCCCGGAGAACCCCTACAAACACCCCGTGT	1933	2954	ACTGGCAGGCCACTTGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGTGTGAAGC	3013
QY	854	TCGCCATCAAGAAGAGGACGACCAAGTGGCGGCAAGCTGTGTGAGCTTCCGCGAGCTGA	913	1934	TGTTGTACCACTGTGAGAGGAGGCCCATTCGCGCGCGGAGACCTTCTAGTGTGACCGCG	1993
Db	1934	TCGCCATCAAGAAGAGGACGACCAAGTGGCGGCAAGCTGTGTGAGCTTCCGCGAGCTGA	1993	3014	TGTTGTACCACTGTGAGAGGAGGCCCATTCGCGCGCGGAGACCTTCTAGTGTGACCGCG	3073
QY	914	ACAAGCCATCCAGAGCTTCTGGAGGTGAGCTGGGCATCCCGCACCCCGCGGCTGA	973	1994	CGCCCAACCGCGAGACCAAGATTCGGCAAGGCCGCTACGTGACCGCGCGGCGCAGA	2053
Db	1994	ACAAGCCATCCAGAGCTTCTGGAGGTGAGCTGGGCATCCCGCACCCCGCGGCTGA	2053	3074	CGCCCAACCGCGAGACCAAGATTCGGCAAGGCCGCTACGTGACCGCGCGGCGCAGA	3133
QY	974	AGAAGAAGAGAGCGTGTGACGCTGGAGCGCGCTACTTCAAGCTGCCCGCTGG	1033	2054	AGATCGTGGAGCTGACCGAGACCCACCAAGAGAGCCGAGCTGACGAGCCATCCAGGTGG	2113
Db	2054	AGAAGAAGAGAGCGTGTGACGCTGGAGCGCGCTACTTCAAGCTGCCCGCTGG	2113	3134	AGATCGTGGAGCTGACCGAGACCCACCAAGAGAGCCGAGCTGACGAGCCATCCAGGTGG	3193
QY	1034	ACGAGGACTTCCGCAAGTACACCGCTTACCATCCCGCAGCATCAACAGAGACCCCG	1093	2114	CCCTCAGAGCAGCGGCGAGCGAGTGAACATCGTGAACGACGAGCGATCCAGGTGGCA	2173
Db	2114	ACGAGGACTTCCGCAAGTACACCGCTTACCATCCCGCAGCATCAACAGAGACCCCG	2173	3194	CCCTCAGAGCAGCGGCGAGCGAGTGAACATCGTGAACGACGAGCGATCCAGGTGGCA	3253
QY	1094	GCATCCGCTACCACTACAACTGTGCTGCCCGAGGGCTGGAAGGGCAGCCCCAGCATCTTC	1153	2174	TCATCCAGGCCAGCCCGACCAAGAGAGCGAGCTGTTGAACAGATCATTCGAGCAGC	2233
Db	2174	GCATCCGCTACCACTACAACTGTGCTGCCCGAGGGCTGGAAGGGCAGCCCCAGCATCTTC	2233	3254	TCATCCAGGCCAGCCCGACCAAGAGAGCGAGCTGTTGAACAGATCATTCGAGCAGC	3313
QY	1154	AGAGCAGATGACCAAGATCTGGAGCGCTTCCGCGCGCGCAACCCCGAGATCGTGTATCT	1213	2234	TGATCAAGAAGAGAGAGTGTACCTGAGCTGGTGCCCGCCCAAGGGCATTCGGCGGCA	2293
Db	2234	AGAGCAGATGACCAAGATCTGGAGCGCTTCCGCGCGCGCAACCCCGAGATCGTGTATCT	2293	3314	TGATCAAGAAGAGAGAGTGTACCTGAGCTGGTGCCCGCCCAAGGGCATTCGGCGGCA	3373
QY	1214	ACCAGGCCCTGTGTACGTGGGACGACCTGGAGATCGGCGCAGCAGCGGCCAAGATCG	1273	2294	ACGAGCAGATTCGACAGCTGTGTGAGCAAGGCGATCCCGCAAGGTGCTGTTCTGGACGCA	2353
Db	2294	ACCAGGCCCTGTGTACGTGGGACGACCTGGAGATCGGCGCAGCAGCGGCCAAGATCG	2353	3374	ACGAGCAGATTCGACAGCTGTGTGAGCAAGGCGATCCCGCAAGGTGCTGTTCTGGACGCA	3433
QY	1274	AGGAGCTGCGAAGACCTGTGTGGCTGGGCTTCAACAACCCCGCAACAGAGCACCAGA	1333	2354	TCGATGGCGGCATCGTGTATCTACCAAGTACATGAGCAGCTGTGAGCGAGCGCGGCC	2413
Db	2354	AGGAGCTGCGAAGACCTGTGTGGCTGGGCTTCAACAACCCCGCAACAGAGCACCAGA	2413	3434	TCGATGGCGGCATCGTGTATCTACCAAGTACATGAGCAGCTGTGAGCGAGCGCGGCC	3493
QY	1334	AGGAGCCCTTCTGCTGCCATCGAGCTGCAACCGCAAGTGGACCGTGCAGGCCATCG	1393	2414	CTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT	2451
Db	2414	AGGAGCCCTTCTGCTGCCATCGAGCTGCAACCGCAAGTGGACCGTGCAGGCCATCG	2473	3494	CTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT	3531
QY	1394	AGCTCCCGAGAGAGAGCTGGACCGTGAACGACATCCAGAGCTGTGTGGCGAGCTGA	1453			
Db	2474	AGCTCCCGAGAGAGAGCTGGACCGTGAACGACATCCAGAGCTGTGTGGCGAGCTGA	2533			

RESULT 14  
US-10-190-435-15  
; Sequence 15, Application US/10190435  
; Publication No. US20030143248A1

GENERAL INFORMATION:  
; APPLICANT: ZUR MEGED, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Yang  
; APPLICANT: ENGELBRSCHT, Susan  
; APPLICANT: VAN KENSBERG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: PP18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 15  
; LENGTH: 3537  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GagPolmutIna\_C  
US-10-190-435-15

Query Match  
Best Local Similarity 99.1%; Score 2402.8; DB 14; Length 3537;  
Matches 2416; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY	14	TGSCCGAGCCATGAGCCAGGCGCACAGCGCCAAACATCTGATGCGAGCGGCAACTTCA	73
DB	1094	TGSCCGAGCGATGAGCCAGGCGCAACACCGCGTGATGATGCGAAGAGCAACTTAA	1153
QY	74	AGGCGCCCAAGCGCATCATCAAGTGTTCATCTGGGCAAGAGGCGCACATCGCCCGCA	133
DB	1154	AGGCGCCCAAGCGCATCATCAAGTGTTCATCTGGGCAAGAGGCGCACATCGCCCGCA	1213
QY	134	ACTGCGCGCCCGCCGAGAGGCGTCTGGAAGTGGCGCAAGAGGCGCACATCGCCCGCA	193
DB	1214	ACTGCGCGCCCGCCGAGAGGCGTCTGGAAGTGGCGCAAGAGGCGCACATCGCCCGCA	1273
QY	194	AGGACTGCAACCGAGCGCCAGGCGCAACTCTTCCGCGAGGAGCTTGCCCTTCCCGAGGCGCA	253
DB	1274	AGGACTGCAACCGAGCGCCAGGCGCAACTCTTCCGCGAGGAGCTTGCCCTTCCCGAGGCGCA	1333
QY	254	AGSCCGCGAGTTCCTCCGAGGAGAGACCGCGCAACAGCGCCCAACAGCGCGAGTGC	313
DB	1334	AGSCCGCGAGTTCCTCCGAGGAGAGACCGCGCAACAGCGCCCAACAGCGCGAGTGC	1393
QY	314	AGGTGCGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	373
DB	1394	AGGTGCGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1453
QY	374	TCCCCAGATACCTGTGTGAGCGCCCTGTTGAGCATCAAGGTGGGCGGCGCAGATCA	433
DB	1454	TCCCCAGATACCTGTGTGAGCGCCCTGTTGAGCATCAAGGTGGGCGGCGCAGATCA	1513
QY	434	AGGAGCGCTGCTGGACACCGCGCGGAGCAGACCGTGTGAGGAGATGAGCGTGGCG	493
DB	1514	AGGAGCGCTGCTGGACACCGCGCGGAGCAGACCGTGTGAGGAGATGAGCGTGGCG	1573
QY	494	GCAAGTGAAGCCCAAGATGATGCGCGCATCGCGGCTTCATCAAGGTGCGCGAGTACG	553
DB	1574	GCAAGTGAAGCCCAAGATGATGCGCGCATCGCGGCTTCATCAAGGTGCGCGAGTACG	1633
QY	554	ACCAGATCTGATCCAGATCTGCGGCAAGAGGCGCATCGGCGCGCGCGCGCGCGCG	613
DB	1634	ACCAGATCTGATCCAGATCTGCGGCAAGAGGCGCATCGGCGCGCGCGCGCGCGCG	1693
QY	614	CCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGTGCGACCTGAACTTC	673
DB	1694	CCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGTGCGACCTGAACTTC	1753
QY	674	CCATCAGCCCCATCGAGACCCGTCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCG	733
DB	1754	CCATCAGCCCCATCGAGACCCGTCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCG	1813

QY	734	TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCGCTTGACCGCCATCTGCGAGGAGA	793
DB	1814	TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCGCTTGACCGCCATCTGCGAGGAGA	1873
QY	794	TGGAGAAGGAGGCGCAAGATCAACCAAGATCGGCGCCCGAGAAACCCCTACAAACACCCCGTGT	853
DB	1874	TGGAGAAGGAGGCGCAAGATCAACCAAGATCGGCGCCCGAGAAACCCCTACAAACACCCCGTGT	1933
QY	854	TGCGCATCAAGAGAGAGACACCAAGTGGCGCGAGCTGGTGGAGCTTCGCGAGCTGA	913
DB	1934	TGCGCATCAAGAGAGAGACACCAAGTGGCGCGAGCTGGTGGAGCTTCGCGAGCTGA	1993
QY	914	ACAAGCGCACCCAGGACTTCTGGAGGTGCAAGTGGGCGATCCCCACCCCGCGCGCTGA	973
DB	1994	ACAAGCGCACCCAGGACTTCTGGAGGTGCAAGTGGGCGATCCCCACCCCGCGCGCTGA	2053
QY	974	AGNAGAAGAGCGTGTGACCTGTGACCTGTGCGCGAGCGCTTACTTACAGGTGCGCCCTGG	1033
DB	2054	AGNAGAAGAGCGTGTGACCTGTGACCTGTGCGCGAGCGCTTACTTACAGGTGCGCCCTGG	2113
QY	1034	ACGAGGACTTCCGCAAGTACACCCGCTTCAACATCCCTCCAGCATCAACAAAGAGACCCCG	1093
DB	2114	ACGAGGACTTCCGCAAGTACACCCGCTTCAACATCCCTCCAGCATCAACAAAGAGACCCCG	2173
QY	1094	GCATCGGTACCACTACCAAGTGTGCGCGAGGCTGGAGGCGAGCGCCAGCATCTTCC	1153
DB	2174	GCATCGGTACCACTACCAAGTGTGCGCGAGGCTGGAGGCGAGCGCCAGCATCTTCC	2233
QY	1154	AGAGCAGCATGACCAAGATCTCTGGAGCGCTTCCCGCGCGCAACCCCGAGATCGTATCT	1213
DB	2234	AGAGCAGCATGACCAAGATCTCTGGAGCGCTTCCCGCGCGCAACCCCGAGATCGTATCT	2293
QY	1214	ACGAGCGCCCTGTACGTGGGCGAGCATCTGGAGATCGGCGAGCACCGCGCGAGATCG	1273
DB	2294	ACGAGCGCCCTGTACGTGGGCGAGCATCTGGAGATCGGCGAGCACCGCGCGAGATCG	2353
QY	1274	AGGAGCTGCGCAAGCACTGTCTGGCGTGGGCGCTTACCAACCCCGCAAGAGCACCAGA	1333
DB	2354	AGGAGCTGCGCAAGCACTGTCTGGCGTGGGCGCTTACCAACCCCGCAAGAGCACCAGA	2413
QY	1334	AGGAGCGCCCTGTCTGGCGATCGAGTGCACCCGCAAGTGGAGACCGTGGAGCGCCATCG	1393
DB	2414	AGGAGCGCCCTGTCTGGCGATCGAGTGCACCCGCAAGTGGAGACCGTGGAGCGCCATCG	2473
QY	1394	AGCTCCCGGAGAGAGAGCTGGAGCGCTGAAACGACATCCAGAGAGCTGGTGGGCAAGCTGA	1453
DB	2474	AGCTCCCGGAGAGAGAGCTGGAGCGCTGAAACGACATCCAGAGAGCTGGTGGGCAAGCTGA	2533
QY	1454	ACTGGGCGCAGCATCTACCCCGCATCAAGGTGCGCGAGCTGTGCAAGCTGTGGCG	1513
DB	2534	ACTGGGCGCAGCATCTACCCCGCATCAAGGTGCGCGAGCTGTGCAAGCTGTGGCG	2593
QY	1514	GCGCCAGCGCCCTGACCGAGATCTGCGCGCTGACCGAGGAGCGAGCTGGAGTGGCG	1573
DB	2594	GCGCCAGCGCCCTGACCGAGATCTGCGCGCTGACCGAGGAGCGAGCTGGAGTGGCG	2653
QY	1574	AGAAACCGGAGAGCTCTGCGCGAGCGCTGACCGCGCTGTACTACGACCCCGAGCAAGACC	1633
DB	2654	AGAAACCGGAGAGCTCTGCGCGAGCGCTGACCGCGCTGTACTACGACCCCGAGCAAGACC	2713
QY	1634	TGGTGGCGGAGATCCAGAGAGCGGCGCACGACGAGTGGAGCTTACAGATCTACAGAGC	1693
DB	2714	TGGTGGCGGAGATCCAGAGAGCGGCGCACGACGAGTGGAGCTTACAGATCTACAGAGC	2773
QY	1694	CTTCAAGAACCTTGAAGACCGGCAAGTACGCGCAAGATGCGCACCGCGCCACACCAAGCAG	1753
DB	2774	CTTCAAGAACCTTGAAGACCGGCAAGTACGCGCAAGATGCGCACCGCGCCACACCAAGCAG	2833
QY	1754	TGAAGCAGCTGACCGAGCGGTGAGAGAGTCCGCGAGGAGCATGTGATCTGGGCA	1813
DB	2834	TGAAGCAGCTGACCGAGCGGTGAGAGAGTCCGCGAGGAGCATGTGATCTGGGCA	2893
QY	1814	AGACCCCAAGTTCGCGCTGCGCATCCAGAGGAGAGCTGGGAGACCTGTGTGGACCGACT	1873

Db 2894 AGACCCCAAGTTCGCGCTGCCCATCCAGAGAGACCTGGAGACCTGGTGGACCGACT 2953  
Qy 1874 ACTGGCAGGACCACTGGATCCCGAGTGGGAGTTCTGTGAACACCCGCCCTGGTGAAGC 1933  
Db 2954 ACTGGCAGGACCACTGGATCCCGAGTGGGAGTTCTGTGAACACCCGCCCTGGTGAAGC 3013  
Qy 1934 TGTGGTACCAAGCTGGAGAGAGCCCATCATCGCGCCGAGACCTTCTAAGTGGACGGCG 1993  
Db 3014 TGTGGTACCAAGCTGGAGAGAGCCCATCATCGCGCCGAGACCTTCTAAGTGGACGGCG 3073  
Qy 1994 CCGCCACCGCGAGACCAAGATCGGACGAGCCGGCTACGTGACCGAGCCGGGCGGCGAGA 2053  
Db 3074 CCGCCACCGCGAGACCAAGATCGGACGAGCCGGCTACGTGACCGAGCCGGGCGGCGAGA 3133  
Qy 2054 AGATCGTGAAGCTGACCGAGACCAACCAACAGAGACCCGAGTGTGACAGGCTCCAGCTGG 2113  
Db 3134 AGATCGTGAAGCTGACCGAGACCAACCAACAGAGACCCGAGTGTGACAGGCTCCAGCTGG 3193  
Qy 2114 CCCTGAGGACAGGCGGAGGAGTGAACATCTGTGACCGACAGCGAGTACGCTGGGCA 2173  
Db 3194 CCCTGAGGACAGGCGGAGGAGTGAACATCTGTGACCGACAGCGAGTACGCTGGGCA 3253  
Qy 2174 TCATCCAGGCGCCAGCCGACAGAGCGAGCGAGTGTGAAACAGATCATCGAGCAGC 2233  
Db 3254 TCATCCAGGCGCCAGCCGACAGAGCGAGCGAGTGTGAAACAGATCATCGAGCAGC 3313  
Qy 2234 TGATCAGAGGAGAGAGTGTACTGTGAGTGGGTGCGGCCACAGAGGAGTCCGGCGCA 2293  
Db 3314 TGATCAGAGGAGAGAGTGTACTGTGAGTGGGTGCGGCCACAGAGGAGTCCGGCGCA 3373  
Qy 2294 ACAGCAGATCGACAACTGGTGTGAGCAAGGGCATCGCAGAGTGTGTTCTTGGACGGCA 2353  
Db 3374 ACAGCAGATCGACAACTGGTGTGAGCAAGGGCATCGCAGAGTGTGTTCTTGGACGGCA 3433  
Qy 2354 TCGATGCGGATCGTGTATCTACAGTACATGAGACAGCTGTGAGTGGGAGCGGCGGCC 2413  
Db 3434 TCGATGCGGATCGTGTATCTACAGTACATGAGACAGCTGTGAGTGGGAGCGGCGGCC 3493  
Qy 2414 CTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451  
Db 3494 CTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 3531

RESULT 15  
US-10-190-435-12  
; Sequence 12, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: P18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 5145  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: GagCompPolmutinatRevNef\_C  
US-10-190-435-12

Query Match 97.8%; Score 2401.8; DB 14; Length 5145;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2403; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 14 TGGCCGAGGCGCATGAGCCAGGCGCACAGCGCCAAACATCTCTGATGCGAGCGAGCAACTTCA 73  
Db 1487 TGGCCGAGGCGCATGAGCCAGGCGCACAGCGCCAAACATCTCTGATGCGAGCGAGCAACTTCA 1546  
Qy 74 AGGCCCCCAAGCGCATCATCAAGTGTCTCAACTCGCGCAAGGAGGCGCACATCGCCCGCA 133  
Db 1547 AGGCCCCCAAGCGCATCATCAAGTGTCTCAACTCGCGCAAGGAGGCGCACATCGCCCGCA 1606  
Qy 134 ACTCGCGCGCCCCCGCAAGAGGGGTCTGGAAGTGGCGCAAGGAGGCGCACAGATGA 193  
Db 1607 ACTCGCGCGCCCCCGCAAGAGGGGTCTGGAAGTGGCGCAAGGAGGCGCACAGATGA 1666  
Qy 194 AGGACTGACCGAGCGCGAGCCCAACTTCTTCCGCGAGGACCTTGGCTTCCCCCAGGGCA 253  
Db 1667 AGGACTGACCGAGCGCGAGCCCAACTTCTTCCGCGAGGACCTTGGCTTCCCCCAGGGCA 1726  
Qy 254 AGGCCCCGAGTTCCTCCAGCGAGCAGAAACCGCGCAACAGCCCAACAGCGCGAGCTGC 313  
Db 1727 AGGCCCCGAGTTCCTCCAGCGAGCAGAAACCGCGCAACAGCCCAACAGCGCGAGCTGC 1786  
Qy 314 AGGTGCGGCGGCAAAACCCCGCAGCGAGCGCGCGCGCGAGCGCGAGGCGACCTGAACT 373  
Db 1787 AGGTGCGGCGGCAAAACCCCGCAGCGAGCGCGCGCGCGCGCGAGGCGACCTGAACT 1846  
Qy 374 TCCCCCAGATCACCCCTGTGCGAGCGCGCCCTGTGTGAGCATCAAGTGGCGCGCGAGATCA 433  
Db 1847 TCCCCCAGATCACCCCTGTGCGAGCGCGCCCTGTGTGAGCATCAAGTGGCGCGCGAGATCA 1906  
Qy 434 AGGAGGCGCTCTCTGAGCAGCGCGCGCGCGCGCGCGCGCGCGAGGAGATGAGCTGCGCG 493  
Db 1907 AGGAGGCGCTCTCTGAGCAGCGCGCGCGCGCGCGCGCGCGCGAGGAGATGAGCTGCGCG 1966  
Qy 494 GCAAGTGAAGCGCCAAAGATGATCGCGCGCATCGCGCGCTTCAACAGTGGCGCGCGAGTACG 553  
Db 1967 GCAAGTGAAGCGCCAAAGATGATCGCGCGCATCGCGCGCTTCAACAGTGGCGCGAGTACG 2026  
Qy 554 ACCAGATCTCTGATCGAGATCTGCGGCAAGAGCGCATCGGACCGTGTGATCGGCGCGCA 613  
Db 2027 ACCAGATCTCTGATCGAGATCTGCGGCAAGAGCGCATCGGACCGTGTGATCGGCGCGCA 2086  
Qy 614 CCGCGTGAACATCATCGCGCGCAACATGCTGACCGAGCTGGGTGACCGCTGAACTTCC 673  
Db 2087 CCGCGTGAACATCATCGCGCGCAACATGCTGACCGAGCTGGGTGACCGCTGAACTTCC 2146  
Qy 674 CCATAGCCCCCATCGAGACCGTGTGCGGAGCTGGAAGCGCGCGGATGAGCGCGCGCGCAAGG 733  
Db 2147 CCATAGCCCCCATCGAGACCGTGTGCGGAGCTGGAAGCGCGCGGATGAGCGCGCGCGCAAGG 2206  
Qy 734 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGCGCTGACCGCGCATCTGCGAGGAGA 793  
Db 2207 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGCGCTGACCGCGCATCTGCGAGGAGA 2266  
Qy 794 TGGAGAGGAGGCGCAAGATCAACAGATCGCGCGCGGAGAACCCCTACAAACCCCGCTGT 853  
Db 2267 TGGAGAGGAGGCGCAAGATCAACAGATCGCGCGCGGAGAACCCCTACAAACCCCGCTGT 2326  
Qy 854 TCGCCATCAAGAGAGGAGCAGCAGCAAGTGGCGCGAGCTGGTGGATTCGCGGAGCTGA 913  
Db 2327 TCGCCATCAAGAGAGGAGCAGCAGCAAGTGGCGCGAGCTGGTGGATTCGCGGAGCTGA 2386  
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAAGCTGGGCGATCCCCCAGCGCGCGCTGA 973  
Db 2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAAGCTGGGCGATCCCCCAGCGCGCGCTGA 2446  
Qy 974 AGAAGAGAGAGCGTGAACCGTGTGAGCTGGGCGAGCGCTACTTCAAGCTGCGCGCTGG 1033  
Db 2447 AGAAGAGAGAGCGTGAACCGTGTGAGCTGGGCGAGCGCTACTTCAAGCTGCGCGCTGG 2506  
Qy 1034 ACGAGGACTTCCGCAAGTACACCGCTTCAACCTCCCGAGCATCAACAGAGACCGCGCG 1093  
Db 2507 ACGAGGACTTCCGCAAGTACACCGCTTCAACCTCCCGAGCATCAACAGAGACCGCGCG 2566

Db	3647	TCATCCAGGCCCGCCGACAGAGCGAGCGAGCTGTGTAAACAGATCATCGAGCAGC	3706
Qy	2234	TGATCAAGAAGGAGAGGTGTACCTGAGCTGGGTGCCGCCCAAGGGGATCGGGGCA	2293
Db	3707	TGATCAAGAAGGAGAGGTGTACCTGAGCTGGGTGCCGCCCAAGGGGATCGGGGCA	3766
Qy	2294	ACGAGCAGATCGACAAGCTGTGTAGCAAGGGCATCCGCAAGGTGCTGTTCTTGGACGGCA	2353
Db	3767	ACGAGCAGATCGACAAGCTGTGTAGCAAGGGCATCCGCAAGGTGCTGTTCTTGGACGGCA	3826
Qy	2354	TCGATGGCGGCATCGTGATCTTACCAAGTACATGACGACCTGTGAGGGAGCGCGCC	2413
Db	3827	TCGATGGCGGCATCGTGATCTTACCAAGTACATGACGACCTGTGAGGGAGCGCGCC	3886
Qy	2414	CTAGS 2418	
Db	3887	CTAGS 3891	

Search completed: April 10, 2004, 21:24:47  
Job time : 597.888 Secs

Qy	1094	GCATCCGCTACCAAGTACAAAGTCTGCTCCCGCAGGGCTGGAAGGGCAGCCCAAGATCTTCC	1153
Db	2567	GCATCCGCTACCAAGTACAAAGTCTGCTCCCGCAGGGCTGGAAGGGCAGCCCAAGATCTTCC	2626
Qy	1154	AGAGCAGATGACCAAGATCTTGGAGCCCTTCGCGCCCGCAACCCCGAGATCGTGATCT	1213
Db	2627	AGAGCAGATGACCAAGATCTTGGAGCCCTTCGCGCCCGCAACCCCGAGATCGTGATCT	2686
Qy	1214	ACCAGGCCCCCTTACGTGGGAGCGACCTGTGAGATCGGCCAGCAACCGCGCAAGATCG	1273
Db	2687	ACCAGGCCCCCTTACGTGGGAGCGACCTGTGAGATCGGCCAGCAACCGCGCAAGATCG	2746
Qy	1274	AGAGCTGGCAGCAACCTTCTGGCTGGGCTTCAACACCCCGCAAGAGACCAAGCA	1333
Db	2747	AGAGCTGGCAGCAACCTTCTGGCTGGGCTTCAACACCCCGCAAGAGACCAAGCA	2806
Qy	1334	AGAGGCCCCCTTCTGGCTGGGCTTCAACACCCCGCAAGAGACCAAGCA	1393
Db	2807	AGAGGCCCCCTTCTGGCTGGGCTTCAACACCCCGCAAGAGACCAAGCA	2866
Qy	1394	AGTGGCCGAGAGGAGATGACCGTGAACGACATTCGAAAGCTGGTGGGCAAGCTGA	1453
Db	2867	AGTGGCCGAGAGGAGATGACCGTGAACGACATTCGAAAGCTGGTGGGCAAGCTGA	2926
Qy	1454	ACTGGGCGAGCAGATCTACCCCGCATCAAGGTGGCCAGCTGTGCAAGCTGCTGGCG	1513
Db	2927	ACTGGGCGAGCAGATCTACCCCGCATCAAGGTGGCCAGCTGTGCAAGCTGCTGGCG	2986
Qy	1514	GGCCAAAGGCCCTGACCGAATCGTCCCGCATCAAGGTGGCCAGCTGTGCAAGCTGCTGGCG	1573
Db	2987	GGCCAAAGGCCCTGACCGAATCGTCCCGCATCAAGGTGGCCAGCTGTGCAAGCTGCTGGCG	3046
Qy	1574	AGAACCGCGAGATCTCGCGGAGCCCGTGACCGCGGTGACTACGACCCCGCAGCAAGGACC	1633
Db	3047	AGAACCGCGAGATCTCGCGGAGCCCGTGACCGCGGTGACTACGACCCCGCAGCAAGGACC	3106
Qy	1634	TGTTGGCCGAGATCCAGAAGCAGGGCCACCAAGTGGACCTTACAGATCTTACAGAGC	1693
Db	3107	TGTTGGCCGAGATCCAGAAGCAGGGCCACCAAGTGGACCTTACAGATCTTACAGAGC	3166
Qy	1694	CTTTCAAGAACCTGAAGACCGGCAAGTACCGCAGATGCGCACCGCCCAACCAAGGACG	1753
Db	3167	CTTTCAAGAACCTGAAGACCGGCAAGTACCGCAGATGCGCACCGCCCAACCAAGGACG	3226
Qy	1754	TGAAGCAGCTGACGAGCCCGTCAGAGATCGCCATGAGAGCATCGTGATCTGGGGCA	1813
Db	3227	TGAAGCAGCTGACGAGCCCGTCAGAGATCGCCATGAGAGCATCGTGATCTGGGGCA	3286
Qy	1814	AGACCCCAAGTTCCGCTGCCATTCAGAGAGGACCTGGGAGACCTGGTGACCGACT	1873
Db	3287	AGACCCCAAGTTCCGCTGCCATTCAGAGAGGACCTGGGAGACCTGGTGACCGACT	3346
Qy	1874	ACTGGCAGGCACTGTGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGTGTAAGC	1933
Db	3347	ACTGGCAGGCACTGTGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGTGTAAGC	3406
Qy	1934	TGTGTACAGCTGGAGAGAGCCCATCTCGCGCGCGAGACCTTCTAGCTGGACGGCG	1993
Db	3407	TGTGTACAGCTGGAGAGAGCCCATCTCGCGCGCGAGACCTTCTAGCTGGACGGCG	3466
Qy	1994	CCGCCAACCGGAGACCAAGATTCGGCAAGCGCGCTACGTGACCGACCGGGGCCGGCAGA	2053
Db	3467	CCGCCAACCGGAGACCAAGATTCGGCAAGCGCGCTACGTGACCGACCGGGGCCGGCAGA	3526
Qy	2054	AGATCGTGAGCTGACGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGG	2113
Db	3527	AGATCGTGAGCTGACGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGG	3586
Qy	2114	CCCTGCAGGACAGCGGAGCGAGGTGAACATCTGTGACCGACAGCCAGTACGCCCTGGGCA	2173
Db	3587	CCCTGCAGGACAGCGGAGCGAGGTGAACATCTGTGACCGACAGCCAGTACGCCCTGGGCA	3646
Qy	2174	TCATCCAGGCCCGCCGACAGAGCGAGCGAGCTGTGTAAACAGATCATCGAGCAGC	2233





QY	398	GGCCCTGGTGAAGATCAAGTGGCGCGCCAGATCAAGGAGCCCTGCTGGACACCGCG	457	Db	1321	GCATCAAGGTGAAGCAGCTGTGCAAGCTGCTCGCGGACCAAGCCCTGACCGAGTGA	1380
Db	241	GGCCCTGGTGAAGATCAAGTGGCGCGCCAGATCAAGGAGCCCTGCTGGACACCGCG	300	QY	1538	TGCCCCCTGACGAGAGCGGAGCTGGAGCTGGCGAGAACCGCGAGATCTCTGGCGGAGC	1597
QY	458	CCGACGACACCGTGTGGAGGAGATGAGCTTCCCGGCAAGTGAAGCCCAAGATGATCG	517	Db	1381	TCCCCCTGACGAGAGCGGAGCTGGAGCTGGCGAGAACCGCGAGATCTCTGAAGGAGC	1440
Db	301	CCGACGACACCGTGTGGAGGAGATGAGCTTCCCGGCAAGTGAAGCCCAAGATGATCG	360	QY	1598	CGGTGCAACCGCTGTACTACGACCCCAAGGACCTGGTGGCCGAGATCTCCGAAGAGCG	1657
QY	518	GGCGATCGCGGCTTCAAGGTGGCGCGAGTACGACCCAGATCTGATCGAGATCTGCG	577	Db	1441	CGGTGCAACCGCTGTACTACGACCCCAAGGACCTGGTGGCCGAGATCTCCGAAGAGCG	1500
Db	361	GGGGATCGGGGCTTCAAGGTGGCGCGAGTACGACCCAGATCTGATCGAGATCTGCG	420	QY	1658	GCACGACGAGTGGACCTTACGAGATCTACGAGAGCCCTTCAAGAACCTGAAGACCGGCA	1717
QY	578	GCAAGAGCCATCGGCAACCGTGTGATCGGCGCCACCCCGTGAACATCATCGGCCGCA	637	Db	1501	GCACGAGCCAGTGGACCTTACGAGATCTACGAGAGCCCTTCAAGAACCTGAAGACCGGCA	1560
Db	421	GCCACAAGGCCATCGGCAACCGTGTGATCGGCGCCACCCCGTGAACATCATCGGCCGCA	480	QY	1718	AGTACCCCAAGATCGGCAACCGCCCAACACGACGCTGAAGCAGTGCACCGAGGCGGTGC	1777
QY	638	ACATCTGACCCAGCTGGGCTGACCCCTGAATCTCCCATCAGCCCATCGAGCCGCTGC	697	Db	1561	AGTACCCCGCATCGCGCGGCCACACCAACGAGCTGAAGCAGTGCACCGAGGCGGTGC	1620
Db	481	ACCTCTGACCCAGATCGGCTGCAACCTGAACTTCCCATCAGCCCATCGAGAGCGGTGC	540	QY	1778	AGAGATCGCCATCGGAGAGATCTGATCTGGGGGCAAGACCCCAAGTTCCGCTGCCCA	1837
QY	698	CGGTGAAGCTGAAGCCCGGATGAGCGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGG	757	Db	1621	AGAGGTGAGCACCGGAGAGCATCTGATCTGGGGCAAGATCCCCAAGTTCAAGCTGCCCA	1680
Db	541	CGGTGAAGCTGAAGCCCGGATGAGCGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGG	600	QY	1838	TCAGAGAGAGACCTTGGGAGACCTGGTGACCGGACTACTTGGCAGGCGCACTTGATCCCGG	1897
QY	758	AGAAGATCAAGCCCTGACCGCATCTGCGAGGATGGAGAGAGGCGGAGATCAACCA	817	Db	1681	TCAGAGAGAGACCTTGGGAGGCTGGTGGATGAGTACTTGGCAGGCGCACTTGATCCCGG	1740
Db	601	AGAAGATCAAGCCCTGCTGGAGATCTGCACCGATGGAGAGGAGGCGGAGATCAACCA	660	QY	1898	AGTGGAGATTCGTGAACACCCCCCTTGGTGAAGCTGTGGTACCGAGCTGGAGAGGAGC	1957
QY	818	AGATCGGCCCCGAGAACCCCTCAACACCCCGCTGTGGCCATCAAGAGAGAGGACAGCA	877	Db	1741	AGTGGAGATTCGTGAACACCCCCCTTGGTGAAGCTGTGGTACCGAGCTGGAGAGAGC	1800
Db	661	AGATCGGCCCCGAGAACCCCTCAACACCCCGCTGTGGCCATCAAGAGAGAGGACAGCA	720	QY	1958	CCATCATCGCGCGGAGACCTTCTACGTGAGAGCGGCGCCCAACCGCGAGACCAAGATCG	2017
QY	878	CCAGTGGCGCAAGCTGTGGTGAATCTCGGAGCTGAACAGCGGCAACCGAGCTTCTGGG	937	Db	1801	CCATCGTGGCGCGGAGACCTTCTACGTGAGAGCGGCGCCCAACCGCGAGACCAAGCTGC	1860
Db	721	CCAGTGGCGCAAGCTGTGGTGAATCTCGGAGCTGAACAGCGGCAACCGAGCTTCTGGG	780	QY	2018	GCAGGCGCGCTACTGTGAACCGGAGCGGCGCGGCGGAGAGTCTGTGAGCTGCACGAGACCA	2077
QY	938	AGTGCAGCTGGGATCCCGCACCCCGCGCTGAAGAGAGAGAGCGTGAACCGTGC	997	Db	1861	GCAGGCGCGCTACTGTGAACCGGAGCGGCGCGGCGGAGAGTCTGTGAGCTGCACGAGACCA	1920
Db	781	AGTGCAGCTGGGATCCCGCACCCCGCGCTGAAGAGAGAGAGCGTGAACCGTGC	840	QY	2078	CCACACGAGAGACCGAGCTGACGGCCATCCAGCTGGCCCTTGCAGGACAGCGGCGAGG	2137
QY	998	TGAGCTGGGCGAGCGCTACTTCAAGCTGCGCTTGAAGAGAGCTTCCGAGATGACCG	1057	Db	1921	CCACACGAGAGACCGAGCTGACGGCCATCCAGCTGGCCCTTGCAGGACAGCGGCGAGG	1980
Db	841	TGAGCTGGGCGAGCGCTACTTCAAGCTGCGCTTGAAGAGAGCTTCCGAGATGACCG	900	QY	2138	TGAACATCGTGAACCGAGCAGTACGCGCTGGGCTATCCAGCTGGGCTATCCAGCTGGGCTATCC	2197
QY	1058	CCCTTACCATCCCGACGATCAACAGAGAGACCCCGGCTACCGTACAGTACCAACGTCG	1117	Db	1981	TGAACATCGTGAACCGAGCAGTACGCGCTGGGCTATCCAGCTGGGCTATCCAGCTGGGCTATCC	2040
Db	901	CCCTTACCATCCCGACGATCAACAGAGAGACCCCGGCTACCGTACAGTACCAACGTCG	960	QY	2198	GCAGAGAGAGCTGGTGAACCGAGATCATCGAGCAGCTGATCAAGAGAGGAGAGGTGTACC	2257
QY	1118	TGCCCCAGGCTGGAGAGGAGCGCCCGAGCATCTTCCAGAGAGAGATGACCAAGATCTGCG	1177	Db	2041	GCAGAGAGAGCTGGTGAACCGAGATCATCGAGCAGCTGATCAAGAGAGGAGAGGTGTACC	2100
Db	961	TGCCCCAGGCTGGAGAGGAGCGCCCGAGCATCTTCCAGAGAGAGATGACCAAGATCTGCG	1020	QY	2258	TGAGCTGGTGGCGCGCCCAAGAGGAGATCGGCGGCAACGAGCAGATCGACAAGCTGGTGA	2317
QY	1178	AGCCCTTCCGCGCCCGACCCCGAGATCGTGAATCTACAGGCGCCCGCTGTACGTGGCA	1237	Db	2101	TGGCTGGTGGCGCGCCCAAGAGGAGATCGGCGGCAACGAGCAGGTGGAGCAAGCTGGTGA	2160
Db	1021	AGCCCTTCCGCGAGAGAACCCCGACATCTGATCTACAGGCGCCCGCTGTACGTGGCA	1080	QY	2318	GCAGGAGATCCGCAAGGTGCTTCTTCTGGAGCGGCTATCGATGGCGGCTATCGTATCTACC	2377
QY	1238	GCAGCTGGAGATCGGCGGACCGCGCAAGATCGAGAGCTGGCGCAAGACCTGTGTC	1297	Db	2161	GCAGGAGATCCGCAAGGTGCTTCTTCTGGAGCGGCTATCGATGGCGGCTATCGTATCTACC	2220
Db	1081	GCAGCTGGAGATCGGCGGACCGCACCAAGATCGAGAGCTGGCGGAGCTGGCGGACCTGCTGC	1140	QY	2378	AGTACATGAGCAGCTGTACGTGGGAGCGGCGGCGCTAGGATCGATTAAGAGCTTCCCG	2437
QY	1298	GCTGGGCTTCAACCCCGGACAGAGACCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG	1357	Db	2221	AGTACATGAGCAGCTGTACGTGGGAGCGGCGGCGCTAGGATCGATTAAGAGCTTCCCG	2280
Db	1141	GCTGGGCTTCAACCCCGGACAGAGACCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG	1200	QY	2438	GGGCTAGCAGCGGTGAATTC	2457
QY	1358	AGTGCACCCCGAGAGTGGAGCGTGCAGGCCATCGAGCTGGCGGAGAGAGAGAGTGA	1417	Db	2281	GGGCTAGCAGCGGTGAATTC	2300
Db	1201	AGTGCACCCCGAGAGTGGAGCGTGCAGGCCATCATGCTGGCCGAGAGAGGAGAGTGA	1260				
QY	1418	CCGTGACGACATCCAGAGCTGGTGGGAGAGCTGAATCTGGGCGAGCGAGATCTACCCG	1477				
Db	1261	CCGTGACGACATCCAGAGCTGGTGGGAGAGCTGAATCTGGGCGAGCGAGATCTACCCG	1320				
QY	1478	GCATCAAGTGGCGGAGCTGTGAAGTGTGGCGGCGCCCAAGGCGCTGACCGACATCG	1537				

RESULT 2  
US-09-475-515-82  
; Sequence 82, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:

```

; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
; LENGTH: 2306
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).Protmod.RTopt.YM
US-09-475-515-82

Query Match      82.2%; Score 2019.2; DB 4; Length 2306;
Best Local Similarity 93.2%; Pred. No. 0;
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

QY 170 GCGGCAAGAGGGCCACCGATGAGGACTGACCGAGGGCCAGGCGGCACTTCTTCGCG 229
DB 1 GCGGCGCGGAGGACACCAATGAAGATTGCACTGAGAGCAGGCTAATTTCTTCGCG 60

QY 230 AGGACTGTCCTTCCCGGCAAGGCGGCGGAGTTCGCCAGCGAGGAGGCGGCGCA 289
DB 61 AGGACTGTCCTTCTGCGGCGGAGGCGGCGGAGTTCAGGAGCGAGGCGGCGCA 120

QY 290 ACAGCGCCACGCGCGGAGTGCAGGTCGAGGTCGCGGCGG-----ACAAAGCGCGCGAGG 343
DB 121 ACAGCGCCACCGCGCGGAGTGCAGGTCGAGGTCGAGGTCGAGTTCAGTTCAGGTCGAGG 180

QY 344 CCGGCGCGGAGCGGCGGAGGCGGCGGCGG-----AATTCGCCAGATCACCTGTGCGAGC 397
DB 181 CCGGCGCGGAGCGGCGGAGGCGGCGGCGG-----AATTCGCCAGATCACCTGTGCGAGC 240

QY 398 GCCCGCTGTGAGCATCAAGTGGCGCGGCGGAGTCAAGAGGCGGCTGTGGAACCGCGG 457
DB 241 GCCCGCTGTGAGCATCAAGTGGCGCGGCGGAGTCAAGAGGCGGCTGTGGAACCGCGG 300

QY 458 CCGAGCACACCGTGTGAGGAGATGAGCTTCCCGGCAAGTGGAGGCCAAGATGATCG 517
DB 301 CCGAGCACACCGTGTGAGGAGATGAGCTTCCCGGCAAGTGGAGGCCAAGATGATCG 360

QY 518 GCGGATCGCGGCTTTCATCAGGTCGCGGAGTACGAGGATCCTGATCGAGATCGG 577
DB 361 GCGGATCGCGGCTTTCATCAGGTCGCGGAGTACGAGGATCCTGATCGAGATCGG 420

QY 578 GCAAGAGGCGCATCGGCGCGGCTGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 637
DB 421 GCAAGAGGCGCATCGGCGCGGCTGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 480

QY 638 ACATGTGACCGAGTGGGCTGACCGCTGAATTCCTCCCATCAGCCCGGCGGCGGCGG 697
DB 481 ACCTGTGACCGAGTGGGCTGACCGCTGAATTCCTCCCATCAGCCCGGCGGCGGCGG 540

QY 698 CCGTGAAGCTGAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 757
DB 541 CCGTGAAGCTGAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600

QY 758 AGAAGATCAAGGCGGCTGACCGCGCATCTGCGAGGAGATGGAAGAGGAGGCGGCAAGATCA 817
DB 601 AGAAGATCAAGGCGGCTGACCGCGCATCTGCGAGGAGATGGAAGAGGAGGCGGCAAGATCA 660

```

```

QY 818 AGATCGGCGCGGAGAACCCCTACAAACACCCCGTGTTCGCCATCAAGAGAGGAGCA 877
DB 661 AGATCGGCGCGGAGAACCCCTACAAACACCCCGTGTTCGCCATCAAGAGAGGAGCA 720

QY 878 CCAAGTGGCGCAAGCTGTGTGAGCTTCGCGGAGCTGAAACAAGCGGACCCAGGACTTCTCG 937
DB 721 CCAAGTGGCGCAAGCTGTGTGAGCTTCGCGGAGCTGAAACAAGCGGACCCAGGACTTCTCG 780

QY 938 AGTGTGAGCTGGGCGATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 997
DB 781 AGTGTGAGCTGGGCGATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840

QY 998 TGGACGTGGGCGAGCGCTACTTTCAGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1057
DB 841 TGGACGTGGGCGAGCGCTACTTTCAGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900

QY 1058 CTTTACACCATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1117
DB 901 CTTTACACCATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960

QY 1118 TGCCCCAGGGCTGGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1177
DB 961 TGCCCCAGGGCTGGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020

QY 1178 AGCCCTTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1237
DB 1021 AGCCCTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080

QY 1238 GCGACCTTGAGAGTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1297
DB 1081 GCGACCTTGAGAGTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140

QY 1298 GCTGGGGCTTACCAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1355
DB 1141 GCTGGGGCTTACCAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200

QY 1356 ----CGAGCTGCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1411
DB 1201 GCTACGAGCTGCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260

QY 1412 GCTGGACCGTGAACGACATCCAGAGCTGTGTGGGCAAGTGAATCTGGGCGGCGGCGGCGG 1471
DB 1261 GCTGGACCGTGAACGACATCCAGAGCTGTGTGGGCAAGTGAATCTGGGCGGCGGCGGCGG 1320

QY 1472 ACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1531
DB 1321 ACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380

QY 1532 ACATCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1591
DB 1381 AGTGTATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440

QY 1592 GCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1651
DB 1441 AGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500

QY 1652 AGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1711
DB 1501 AGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560

QY 1712 CCGGCAAGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1771
DB 1561 CCGGCAAGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620

QY 1772 CCGTGCAGAGATCGCCATGAGGAGGATCGTGTCTGGGCGGCGGCGGCGGCGGCGGCGG 1831
DB 1621 CCGTGCAGAGATCGCCATGAGGAGGATCGTGTCTGGGCGGCGGCGGCGGCGGCGGCGG 1680

QY 1832 TGCCCATCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1891
DB 1681 TGCCCATCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1740

QY 1892 TCCCGGAGTGGGAGTTCGTGAACACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1951

```

```
Db 1741 TCCCGGAGTGGAGTTCTGTGAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGGAGA 1800
Qy 1952 AGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGAAGCGCGCCGCCAACCGGAGACCA 2011
Db 1801 AGGAGCCCATCGTGGCGCGAGACCTTCTACGTGGAAGCGCGCCGCCAACCGGAGACCA 1860
Qy 2012 AGATCGGCAAGCCGCGTACGTGACCAACCGGCGCGGCGGAGAGATCGTAGCCTGACCG 2071
Db 1861 AGCTGGGCAAGCCGCGGTACGTGACCAACCGGCGCGGCGGAGAGGTGGTAGCATCGCCG 1920
Qy 2072 AGACCAACCAACAGAGACCGAGCTGACAGGACATCCAGCTGGCCCTGACAGGACAGGGCA 2131
Db 1921 ACACCAACCAACAGAGACCGAGCTGACAGGACATCCACCTGGCCCTGACAGGACAGGGCC 1980
Qy 2132 GCGAGGTGAACATCGTGAACCAAGCCAGTACAGGACCTGGGCAATCATCCAGGCCAGCCCG 2191
Db 1981 TGGAGGTGAACATCGTGAACCAAGCCAGTACAGGACCTGGGCAATCATCCAGGCCAGCCCG 2040
Qy 2192 ACAAGAGCGAGCGAGCTGTGTAACAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2251
Db 2041 ACAAGAGCGAGCGAGCTGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100
Qy 2252 TGTACCTGAGTGGGTGCGCCGCCCAAGAGGCAATCGGCGGCAACGAGCAGATCGAAGC 2311
Db 2101 TGTACCTGGCTGGGTGCGCCGCCCAAGAGGCAATCGGCGGCAACGAGCAGTGGACAAGC 2160
Qy 2312 TGCTGAGCAAGCGATCGCAAGTGTGTTCTTGGAGCGGATCGATCGGCGGATCGTGA 2371
Db 2161 TGCTGAGCGCGGATCGCGAAGTGTGTTCTTGAACGCGATCGATGGCGGATCGTGA 2220
Qy 2372 TCTACCACTATGAGCAACCTGTACGTGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2431
Db 2221 TCTACCACTATGAGCAACCTGTACGTGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2280
Qy 2432 TTCCCGGGCTAGCACCGGTGAATTC 2457
Db 2281 TTCCCGGGCTAGCACCGGTGAATTC 2306
```

## RESULT 3

```
US-09-475-515-84
; Sequence 84, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 84
; LENGTH: 2312
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.Rtopt(+)
US-09-475-515-84
```

```
Query Match 81.3%; Score 1998.4; DB 4; Length 2312;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;
```

```
Qy 170 GCGCAGGAGGGCCACCGATGAGACTGCAACCGAGCGCCAGGCCAACTTCTTCGCG 229
Db 1 GCGCGCGGAGAGGACACCAATGAAGATTGCTATGAGACAGGCTTAATTTCTTCGCG 60
Qy 230 AGGACCTGCGCTTCCCGCAGGCAAGCGCCGAGTTCCCGAGGAGAGAAACCGCGCCA 289
Db 61 AGGACCTGCGCTTCCCGCAGGCAAGCGCCGAGTTCAAGCAGCAGCAGACCGCGCCA 120
Qy 290 ACAGCCCCACAGCGCGGAGCTGAGGTGGCGGCG-----ACAAACCCCGCAGCGAGG 343
Db 121 ACAGCCCCACCGCGCGGAGCTGAGGTGGCGGCGGAGAGAACACAGCTGAGCGAGG 180
Qy 344 CCGGCGCGAGCGCGCAGGCGCACCTG-----AACTTCCCGCAGATCACTGTGTGAGC 397
Db 181 CCGGCGCGACCGCGCAGGCGCACCTGAGCTTCAACTTCCCGCAGATCACTGTGTGAGC 240
Qy 398 GCCCCTGAGCATCAAGGTGGCGCGCAGATCAAGGAGGCGCTTGTGACACCGCG 457
Db 241 GCCCCTGAGCATCAAGGTGGCGCGCAGCTCAAGGAGGCGCTTGTGACACCGCG 300
Qy 458 CCGCAGCACCGTGTGAGGAGATGAGCTTCCCGCAAGTGAAGCCCAAGATGATCG 517
Db 301 CCGCAGCACCGTGTGAGGAGATGAGCTTCCCGCAAGTGAAGCCCAAGATGATCG 360
Qy 518 GCGCATCGCGGCTTCAATCAAGGTGGCGCAGTACGAGCAGATCTGATCGAGATCTGCG 577
Db 361 GCGCATCGCGGCTTCAATCAAGGTGGCGCAGTACGAGCAGATCTGATCGAGATCTGCG 420
Qy 578 GCAAGAGGCCATCGGCAACCGTGTGATCGGCGCGCCACCGCGTGAACATCATCGCGCGCA 637
Db 421 GCAAGAGGCCATCGGCAACCGTGTGATCGGCGCGCGCCACCGCGTGAACATCATCGCGCGCA 480
Qy 638 ACATGTCAGCAGCTGGGCTGACCTGAACCTTCCCGCATCAGCCCCATCAGAGCCGTGC 697
Db 481 ACCTGTCAGCAGCTGGGCTGACCTGAACCTTCCCGCATCAGCCCCATCAGAGCCGTGC 540
Qy 698 CCGTGAAGCTGAAGCCCGCGCATGGAGCGCCCAAGTGAAGCAGTGGCGCTTGAACCGAGG 757
Db 541 CCGTGAAGCTGAAGCCCGCGCATGGAGCGCCCAAGTGAAGCAGTGGCGCTTGAACCGAGG 600
Qy 758 AGAAGATCAAGGCCCTGACCGCATCTCGAGGAGATGGAGAGAGGCGGCAAGATCACA 817
Db 601 AGAAGATCAAGGCCCTGAGTGAAGTCTGACCGAGATGGAGAGAGGCGGCAAGATCACA 660
Qy 818 AGATCGCGCGGAGAACCCCTACACACCCCGCTGTCGCCATCAAGAGAGAGGACAGCA 877
Db 661 AGATCGCGCGGAGAACCCCTACACACCCCGCTGTCGCCATCAAGAGAGAGGACAGCA 720
Qy 878 CCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAAACAAAGCGCACCCAGGACTTCTGG 937
Db 721 CCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAAACAAAGCGCACCCAGGACTTCTGG 780
Qy 938 AGTGAAGTGGCGCATCCCCACCGCGCGCTCAAGAGAGAGAGCGGTGACCGTGC 997
Db 781 AGTGAAGTGGCGCATCCCCACCGCGCGCTCAAGAGAGAGAGCGGTGACCGTGC 840
Qy 998 TGGACGTGGCGACCGCTACTTTCAGCGTGCCTCGAGCAGGACTTTCGCGAAGTACACCG 1057
Db 841 TGGACGTGGCGACCGCTACTTTCAGCGTGCCTCGAGCAGGACTTTCGCGAAGTACACCG 900
Qy 1058 CTTTCAACATCCCGCAGCATCAACAGGAGACCCCGCGATCCGCTACCACTACACGTCG 1117
Db 901 CTTTCAACATCCCGCAGCATCAACAGGAGACCCCGCGATCCGCTACCACTACACGTCG 960
Qy 1118 TGCCCCAGGGCTGGAAGGGCAGCCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCGG 1177
Db 961 TGCCCCAGGGCTGGAAGGGCAGCCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCGG 1020
Qy 1178 AGCCCTTCGCGCGCGCCCAACCCCGAGATCGTGTACTACCA-----GGCCCCCTGTAGC 1231
Db 1021 AGCCCTTCGCGAGAGCAACCCCGAGATCGTGTACTACCACTACAGTACATGGACGACCTGTACG 1080
```



QY 419 TGGCGCGCCAGATCAAGAGAGCCCTCTCTGGAACAACCGCGCCGACGACACCGTGTCTGAGG 478  
Db 1519 TGGCGCGCCAGATCAAGAGAGCCCTCTCTGGAACAACCGCGCCGACGACACCGTGTCTGAGG 1578  
QY 479 AGATGAGCTGCCCGGCAAGTGAAGCCCAAGATGATCGCGGATCGCGGCTTCATCA 538  
Db 1579 AGATGAGCTGCCCGGCAAGTGAAGCCCAAGATGATCGCGGATCGCGGCTTCATCA 1638  
QY 539 AGGTGGCCAGTACGACCAAGATCCTGATCGAGATCTGGGCAAGAGGCAATCGGACCG 598  
Db 1639 AGGTGGCCAGTACGACCAAGATCCTGATCGAGATCTGGGCAAGAGGCAATCGGACCG 1698  
QY 599 TGTGATCGGCGCCCAAGTGAAGCCCAAGATGATCGCGGATCGCGGCTTCATCA 658  
Db 1699 TGTGATCGGCGCCCAAGTGAAGCCCAAGATGATCGCGGATCGCGGCTTCATCA 1758  
QY 659 GCACCTGAACTTCCCATCAGCCCATCGAGCCGATCGAGCCGATCGAGCCGATCGAGCCG 718  
Db 1759 GCACCTGAACTTCCCATCAGCCCATCGAGCCGATCGAGCCGATCGAGCCGATCGAGCCG 1818  
QY 719 TGAAGCGCCCAAGTGAAGCCCAAGATGATCGCGGATCGCGGCTTCATCA 778  
Db 1819 TGAAGCGCCCAAGTGAAGCCCAAGATGATCGCGGATCGCGGCTTCATCA 1878  
QY 779 CCATCTCGAGGAGATGGAAGAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 838  
Db 1879 AGATCTGACCGAGATGGAAGAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 1938  
QY 839 ACAACACCCCGTGTTCGATCAAGAGAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 898  
Db 1939 ACAACACCCCGTGTTCGATCAAGAGAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 1998  
QY 899 ACTTCCGCGAGCTGAAACAGCGCAACCGAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 958  
Db 1999 ACTTCCGCGAGCTGAAACAGCGCAACCGAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 2058  
QY 959 ACCCGCGCGCTGAAAGAGAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 1018  
Db 2059 ACCCGCGCGCTGAAAGAGAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 2118  
QY 1019 TCAGCGTCCCTGAGAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 1078  
Db 2119 TCAGCGTCCCTGAGAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 2178  
QY 1079 ACAACGAGACCCCGGATCGGATCAAGAGAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 1138  
Db 2179 ACAACGAGACCCCGGATCGGATCAAGAGAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 2238  
QY 1139 GCGCGAGCATCTCCAGAGCAGATGACCAAGATCTGAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 1198  
Db 2239 GCGCGAGCATCTCCAGAGCAGATGACCAAGATCTGAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 2298  
QY 1199 CCGAGATCGTATCAACA-----GGCGCGCGCTGATCGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 1252  
Db 2299 CCGAGATCGTATCAACA-----GGCGCGCGCTGATCGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 2358  
QY 1253 GCGAGCAGCGCGGAGATCGAGAGGCTGCGCAAGACCTGCTGCGTGGGCTTCACCA 1312  
Db 2359 GCGAGCAGCGCGGAGATCGAGAGGCTGCGCAAGACCTGCTGCGTGGGCTTCACCA 2418  
QY 1313 CCGCGAGCAAGAGCAGAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 1366  
Db 2419 CCGCGAGCAAGAGCAGAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 2478  
QY 1367 CCGAGAGTGAAGCGTGAAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 1426  
Db 2479 CCGAGAGTGAAGCGTGAAGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 2538  
QY 1427 ACATCCAGAGTGTGGGCAAGTGAAGTGGGCGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 1486  
Db 2539 ACATCCAGAGTGTGGGCAAGTGAAGTGGGCGGAGGCAAGATCAACAGATCGGCGCCGAGAACCCCT 2598

QY 1487 TGGCGAGCTGTGCAAGCTCTGCGGCGCCCAAGGCGCTTACCGACATCTGTCGCCCTGA 1546  
Db 2599 TGAAGCAGCTGTGCAAGCTCTGCGGCGCCCAAGGCGCTTACCGAGGTGATCCCGCTGA 2658  
QY 1547 CCGAGGAGCGGAGCTGAGAGCTGCGGAGAAACCGCAGATCTCTGCGGAGCCCGTGCAGC 1606  
Db 2659 CCGAGGAGCGGAGCTGAGAGCTGCGGAGAAACCGCAGATCTCTGAAAGGAGCCCGTGCAGC 2718  
QY 1607 GCGGTACTACGACCCAGCAAGAGCTGCTGTCGCGAGATCTCCAGAAAGCAGGCGCAAGAAC 1666  
Db 2719 AGGTGTACTACGACCCAGCAAGAGCTGCTGTCGCGAGATCTCCAGAAAGCAGGCGCAAGAAC 2778  
QY 1667 AGTGGACCTACAGATCTACAGAGGCGCTTCAAGAACTTGAAGACCGGCAAGTACGCCA 1726  
Db 2779 AGTGGACCTACAGATCTACAGAGGCGCTTCAAGAACTTGAAGACCGGCAAGTACGCCA 2838  
QY 1727 AGATGGCAACCGCCCAACCAAGCTGAGAGCTGACCGAGGCGCTGCAAGAGATCG 1786  
Db 2839 GCATGGCGGCGCCCAACCAAGCTGAGAGCTGACCGAGGCGCTGCAAGAGATCG 2898  
QY 1787 CCATGGAGAGCATCTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGG 1846  
Db 2899 GCACGAGAGCATCTGATCTGGGGCAAGATCTCCCAAGTTCCAGCTGCCCATCCAGAGG 2958  
QY 1847 AGATGGAGACCTGCTGAGACCGACTACTGGGAGGCGCACTGGATCCCGAGTGGGAGT 1906  
Db 2959 AGATGGAGACCTGCTGAGACCGACTACTGGGAGGCGCACTGGATCCCGAGTGGGAGT 3018  
QY 1907 TCCTGAAACACCCCGCTGCTGAGAGCTGTGTACCAAGTGGAGAGGAGCCCATCATCG 1966  
Db 3019 TCCTGAAACACCCCGCTGCTGAGAGCTGTGTACCAAGTGGAGAGGAGCCCATCATCG 3078  
QY 1967 GCGCGAGACCTTCTAGCTGAGCGCGCGCCCAACCGGAGACCAAGATCGGCAAGCGC 2026  
Db 3079 GCGCGAGACCTTCTAGCTGAGCGCGCGCCCAACCGGAGACCAAGATCGGCAAGCGC 3138  
QY 2027 GCTAGTACCGACCGAGGCGCGGAGAGATCTGAGCTTACCGAGACCAACCAACAG 2086  
Db 3139 GCTAGTACCGACCGAGGCGCGGAGAGATCTGAGCTTACCGAGACCAACCAACAG 3198  
QY 2087 AGACGAGCTGAGGCGCATCCAGCTGCGCTGCGAGGAGCGGAGGAGTGAACATCG 2146  
Db 3199 AGACGAGCTGAGGCGCATCCAGCTGCGCTGCGAGGAGCGGAGGAGTGAACATCG 3258  
QY 2147 TGACCGACGAGCAGTACCGCTTGGGCTATCTCCAGGCGCCAGCCCGAGAGAGCG 2206  
Db 3259 TGACCGACGAGCAGTACCGCTTGGGCTATCTCCAGGCGCCAGCCCGAGAGAGCG 3318  
QY 2207 AGCTGTTGAAACAGATCTAGAGAGCTGATCAAGAGGAGAGGAGTGTACCTGAGCTGG 2266  
Db 3319 AGCTGTTGAGCGAGATCTAGAGAGCTGATCAAGAGGAGAGGAGTGTACCTGAGCTGG 3378  
QY 2267 TGCGCGCGCCCAAGGCGCATCGCGCAACGAGCAGATCGCAAGCTGTGAGCAAGGCA 2326  
Db 3379 TGCGCGCGCCCAAGGCGCATCGCGCAACGAGCAGGAGTGAACAGCTGTGAGCGCGCA 3438  
QY 2327 TCCGCAAGGAGTGTCTGAGCGGCTGAGTGGCGGATCTGATCTTACAGTA 2381  
Db 3439 TCCGCAAGGAGTGTCTGAGCGGCTGAGTGGCGGATCTGATCTTACAGTA 3493

## RESULT 5

US-09-475-515-81  
; Sequence 81, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine

APPLICANT: SELBY, Mark  
APPLICANT: WALKER, Christopher  
TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
FILE REFERENCE: 1621.002  
CURRENT APPLICATION NUMBER: US/09/475.515A  
CURRENT FILING DATE: 1999-12-30  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 81  
LENGTH: 2299  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE: Description of Artificial Sequence:  
OTHER INFORMATION: FS(+).prolnact.R10pt.YMMW  
US-09-475-515-81

Query Match 76.2%; Score 1872.8; DB 4; Length 2299;  
Best Local Similarity 89.3%; Pred. No. 9.6e-305;  
Matches 2055; Conservative 0; Mismatches 232; Indels 13; Gaps 3;

Qy 170 GCGCAAGGAGGCGCACAGATGAAGACTGACCCGAGCGCCAGGCCAACTTCTTCGGCG 229  
Db 1 GCGGCGCGAAGGACACCAATGAAGATTGCACTGAGACAGGCTTAATTTTT-AGGG 59

Qy 230 AGGACCTGGCTTCCCCAGGCGCAAGCCCGAGTTTCCCCAGCGAGCAGAACCGCGCCA 289  
Db 60 AAGATCTGGCTTCTTACAGGGAAGCCAGGGAAATTTCTTCAGAGCAGACAGGCCA 119

Qy 290 ACAGCCCCACAGCGCGGAGTGTGAGTGGCGCGCG-----ACAACCCCGCAGCGAGG 343  
Db 120 ACAGCCCCACAGGAAGAGAGCTTCAGGTTTGGGAGGAGAAACAACTCCCTCTCAGAAG 179

Qy 344 CCGCGCGCGCGCCAGGCG- ----CCCTGACTTCCCCAGATCACCTGTGGCAGC 397  
Db 180 CAGAGCGGATAGAACAGAACTGTATCTTTAACTTCCCTCAGATCACTTTTGGCAAC 239

Qy 398 GCCCCTTGTGAGCATCAAGTGGCGGCCAGATCAAGGAGGCGCTCTGGACACCGCG 457  
Db 240 GACCCCTGTCAATAGGATCGGGGGCAACTCAAGGAAGCGCTCTGTATACAGAG 299

Qy 458 CCGAGCACACGCTGTGAGAGATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATCG 517  
Db 300 CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAATGGAACCAAAAATGATAG 359

Qy 518 CCGGATCGGCGGCTTCAATCAAGTGGCCAGTACAGACAGATCCTGATCGAGATCTGG 577  
Db 360 GGGGATCGGGGCTTCAATCAAGTGGCCAGTACAGACAGATACCTGTAGAAATCTGTG 419

Qy 578 GCAAGAGGCCATCGGACCGTGTGATCGGCGCCACCCCGTGAAATCATCGGCGCA 637  
Db 420 GACATAAGCTTATAGTACAGTATTAGTAGACCTACACCTGTCAAATANTTGAAGAA 479

Qy 638 ACATGCTGACCCAGCTGGGCTGCACTTGAATTTCCCATCAGCCCATCGAGACCGTGC 697  
Db 480 ATCTTTGACCCAGATCGGCTGCACTTGAATTTCCCATCAGCCCATCGAGACCGTGC 539

Qy 698 CCGTGAAGCTGAGCCCGGATGAGCGCCCAAGTGAAGTGGCGCCCTGACCGAGG 757  
Db 540 CCGTGAAGTGAAGCCCGGATGAGCGCCCAAGTGAAGTGGCGCCATGAGCCAGG 599

Qy 758 AGAAGATCAAGGCCCTGACCGCATCTCGAGGAGATGGAAGAGGCGGCAAGATCACCA 817  
Db 600 AGAAGATCAAGGCCCTGTGTGAGATCTGCAACGAGATGGAGAGGCGCAAGATCAGCA 659

Qy 818 AGATCGGCGCGAGAACCCCTACACCGCCGTTGTCCTCATCAAGGAAGAGCAGCA 877  
Db 660 AGATCGGCGCGAGAACCCCTACACACCCCGTTTCCGCTCATCAAGGAAGAGCAGCA 719

Qy 878 CCAAGTGGCGCAAGTGTGTGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937  
Db 720 CCAAGTGGCGCAAGTGTGTGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 779

Qy 938 AGGTGAGCTGGGATCCCCCAACCCCGCGGCTTGAAGAAAGAGAGCGTGAACGCTGC 997  
Db 780 AGGTGAGCTGGGATCCCCCAACCCCGCGGCTTGAAGAAAGAGAGCGTGAACGCTGC 839

Qy 998 TGGACGTGGGCGAGCGCTTACTTTCAGCGTGCCTTGGACGAGACTTCCGCAAGTACACCG 1057  
Db 840 TGGACGTGGGCGAGCGCTTACTTTCAGCGTGCCTTGGACGAGACTTCCGCAAGTACACCG 899

Qy 1058 CTTTCAACCATCCGAGCATCAACAAAGAGACCCCGGATCCGCTACAGTACACGCTGC 1117  
Db 900 CTTTCAACCATCCGAGCATCAACAAAGAGACCCCGGATCCGCTACAGTACACGCTGC 959

Qy 1118 TGCCCCAGGCTTGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGG 1177  
Db 960 TGCCCCAGGCTTGAAGGCGAGCCCGGAGCTTCCAGAGCAGCATGACCAAGATCTCTGG 1019

Qy 1178 AGCCCTTCCGCGCGCAACCCCGAGATCTGTGATCTTACAGGCGCCCTGTACGTGGCA 1237  
Db 1020 AGCCCTTCCGCGCAAGCAGAAACCCCGAGCATCTGTGATCTTACAGGCGCCCTGTACGTGGCA 1079

Qy 1238 GCGACCTGGAGATGCGCCAGCACCGCGCAAGATCGAGAGCTGCGCAAGCAGCTGTCTGC 1297  
Db 1080 GCGACCTGGAGATGCGCCAGCACCGCACCAAGATCGAGAGCTGCGCCAGCAGCTGTCTGC 1139

Qy 1298 GCTGGGGCTTCAACACCCCGACAAAGAGCACCAGAGAGGCGCCCTTCTTGCCTGC 1357  
Db 1140 GCTGGGGCTTCAACACCCCGACAAAGAGCACCAGAGAGGCGCCCTTCTTGCCTGC 1199

Qy 1358 AGCTGCAACCCGACAAAGTGGACCGTGCAGCGCCATCGAGCTGCCGAGAGAGAGAGCTGGA 1417  
Db 1200 AGCTGCAACCCGACAAAGTGGACCGTGCAGCGCCATCGAGCTGCCGAGAGAGAGAGCTGGA 1259

Qy 1418 CCGTGAAACGACATCCAGAAAGCTGTGGCAAGCTGGAAGTGGGCGAGCCAGATCTTACCCCG 1477  
Db 1260 CCGTGAAACGACATCCAGAAAGCTGTGGCAAGCTGGAAGTGGGCGAGCCAGATCTTACCCCG 1319

Qy 1478 GATCAAGTGGCGCAGCTGTGCAAGCTGCTGGCGCGCCCAAGCCCTGACCGAGCTGC 1537  
Db 1320 GATCAAGTGAAGCAGCTGTGCAAGCTGCTGGCGCGCCCAAGCCCTGACCGAGCTGA 1379

Qy 1538 TGCCTCTGACCGAGAGGCGGAGCTGGAGCTGGCGGAGAACCGCGAGATCTTGCOCGAGC 1597  
Db 1380 TGCCTCTGACCGAGAGGCGGAGCTGGAGCTGGCGGAGAACCGCGAGATCTTGAAGGAGC 1439

Qy 1598 CCGTGCAACCGCTGTACTACGACCCCGCAGCAAGACTTGGTGGCGGAGATCCAGAAAGCAGG 1657  
Db 1440 CCGTGCAACCGCTGTACTACGACCCCGCAGCAAGACTTGGTGGCGGAGATCCAGAAAGCAGG 1499

Qy 1658 GCGACGACGAGTGGACCTTACAGATCTACAGAGGCGCTTCAAGAACTGAGACCGGCA 1717  
Db 1500 GCGAGGCGCAGTGGACCTTACAGATCTACAGAGGCGCTTCAAGAACTGAGACCGGCA 1559

Qy 1718 AGTACGCCAAGATGCGCACCGCCCAACCAACGAGCTGAAGCAGCTGACCGAGGCGCTGC 1777  
Db 1560 AGTACGCCCGCATGCGCGCGCCCAACCAACGAGCTGAAGCAGCTGACCGAGGCGCTGC 1619

Qy 1778 AGAAGATCGCCATGGAGAGCATCTGTGATCTGGGCGAAGACCCCAAGTTCGCCCTGCCCA 1837  
Db 1620 AGAAGTGAAGCAGCAGAGCATCTGTGATCTGGGCGAAGATCCCAAGTTCAGAGCTGCCCA 1679

Qy 1838 TCCAGAGAGAGACCTTGGAGAGCTTGGTGGACCGACTTCTGGCAGGCGCCCTGATCCCG 1897  
Db 1680 TCCAGAGAGAGACCTTGGAGAGCTTGGTGGATGAGTACTGGCAGGCGCCCTGATCCCG 1739

Qy 1898 AGTGGGAGTTCTGTAAACACCCCGCCCTTCTGTGAAGCTGTGTAACAGCTGGAGAGAGC 1957  
Db 1740 AGTGGGAGTTCTGTAAACACCCCGCCCTTCTGTGAAGCTGTGTAACAGCTGGAGAGAGC 1799

Qy 1958 CCAATCGCGCGCGAGACCTTCTAGTGGACGCGCGCCCAAGCCGAGAGACCAAGATCG 2017  
Db 1800 CCAATCGTGGCGCGGAGACCTTCTACGTGGACGCGCGCCCAAGCCGAGAGACCAAGCTCG 1859



```

Qy 2018 GCAAGGCGGCTACGTGACCGACCGGGCGCGCAGAGAATCGTGAGCTGACCGAGACCA 2077
Db 1860 GCAAGGCGGCTACGTGACCGACCGGGCGCGCAGAGAATCGTGAGCTGACCGAGACCA 1919
Qy 2078 CCACACGAGACCGAGCTGAGGCGCATCCAGCTGGCCCTGCGAGGACGCGGCGAGG 2137
Db 1920 CCAACACGAGACCGAGCTGAGGCGCATCCAGCTGGCCCTGCGAGGACGCGGCTGAGG 1979
Qy 2138 TGAACATCGTGACCGACAGCGACGTACGCGCTGGGCGCATCCAGGCGCGACGACGA 2197
Db 1980 TGAACATCGTGACCGACAGCGACGTACGCGCTGGGCGCATCCAGGCGCGACGACGA 2039
Qy 2198 GCGAGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGACC 2257
Db 2040 GCGAGAGCGAGCTGGTGAAGCGAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGACC 2099
Qy 2258 TGAGCTGGTGCGCGCGCCACAAAGGCGCATCGCGCGCAACGAGCAGATCGAACGCTGGTGA 2317
Db 2100 TGGCCTGGTGCGCGCGCCACAAAGGCGCATCGCGCGCAACGAGGAGGTGGAACGCTGGTGA 2159
Qy 2318 GCAAGGCGATCCGCAAGGTGCTGTCCTGACGCGCATCGATGCGGCGCATCGTGTATACC 2377
Db 2160 GCGCGCGCATCCGCAAGGTGCTGTCCTGACGCGCATCGATGCGGCGCATCGTGTATACC 2219
Qy 2378 AGTACATGGAGCACTGTACGTGGGCGAGCGCGCGCCCTAGGATCGATTAAAGCTTCCCG 2437
Db 2220 AGTACATGGAGCACTGTACGTGGGCGAGCGCGCGCCCTAGGATCGATTAAAGCTTCCCG 2279
Qy 2438 GGGCTAGCACCGGTGAATTC 2457
Db 2280 GGGCTAGCACCGGTGAATTC 2299

```

RESULT 6

```

US-09-475-515-80
; Sequence 80, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
; LENGTH: 2305
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(+).proinact.Rtopt.YM
US-09-475-515-80

```

```

Query Match 75.4%; Score 1852; DB 4; Length 2305;
Best Local Similarity 89.0%; Pred. No. 2.9e-301;
Matches 2052; Conservative 0; Mismatches 235; Indels 19; Gaps 4;
Qy 170 GCGGCAAGGAGGCGCCACAGAGTGAAGGAGTGCACCGCGCGCAGGCGCAACTTCTTCGCGC 229
Db 1 GCGGCGCGGAGGACACCAATGAAGATTGCACCTGAGAGACAGCTTAATTTTT-AGGG 59
Qy 230 AGGACTGGCTTCCCGGCGGAGGCGCGGAGTTCGCCGCGAGGCGCGGCGCA 289

```

```

Db 60 AAGATCTGGGCTTCTCTCAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACACAGAGCCA 1119
Qy 290 ACAGCCCCACAGCGCGAGCTGCAGGTGCGGCGG-----ACAAACCCCGCAGGAGG 343
Db 120 ACAGCCCCACAGGAAGAGGCTTCAGGTTTGGGAGGAGAAACAACTCCCTCTCAGAAG 179
Qy 344 CCGGCGCGGAGCGCGCAGGGA-----CCCTGACTTCCCGCAGATCACCTGTGGCAGC 397
Db 180 CAGGAGCGGATAGACAAGGAACCTGTATCTTTAACTTCCCTCAGATCCTCTTTTGGGAAC 239
Qy 398 GCCCCTCTGATGAGCATCAAGGTGGGCGCGCAGATCAAGAGGAGGCGCCTCTCGACACCGCG 457
Db 240 GACCCCTCTGTCACAAATAGGATCGGGGGAACCTCAAGGAAGCGCTGCTCGATACAGAG 299
Qy 458 CCGAGCAGCACCGTCTGAGGAGATGAGCTGCGCGCAAGTGAAGCCGCAAGATGATCG 517
Db 300 CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAATGGAACCAACCAAAATGATAG 359
Qy 518 GCGGCATCGCGGCTTCAATCAAGGTGGCGCAGTACGACCAAGATCCTGATCGAGATCTGCG 577
Db 360 GGGGATCGGGGCTTCAATCAAGGTGAGGAGTACGACCAAGATACCTGTAGAAATCTGTG 419
Qy 578 GCAAGAGCGCATCGGCAACCGTGTGATCGGCGCCACCGCGTGAAATCATCATCGCGCGCA 637
Db 420 GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAATCAATTTGGAAGAA 479
Qy 638 ACATGCTGACCGAGCTGGGCTGCACCTGAACTTCCCGCATCAGCGCCCATCGAGACCGTGC 697
Db 480 ATCTGTTGACCGAGATCGGCTGCACTTGACCTTCCCGCATCAGCGCTATTGAGACGGTGC 539
Qy 698 CCGTGAAGCTGAAGCGCGCATGGAAGCGCCCAAGGTGAAGCAGTGGCCCTGAGCCAGG 757
Db 540 CCGTGAAGTTGAAGCGCGGATGGAAGCGCCCAAGGTCAAGCAATGGCCATTGACCGAGG 599
Qy 758 AGAAGATCAAGCGCTGACCGCATCTGCGAGGAGATGGAAGAGGAGGCGGAAGATCAACCA 817
Db 600 AGAAGATCAAGCGCTTGGTGAGATCTGACCGAGATGGAAGAGGAGGCGGCAAGTACGCA 659
Qy 818 AGATCGCGCGCGAGAACCCCTACAAACCCCGCTGTCGCATCAAGAAAGAGACAGCA 877
Db 660 AGATCGCGCGCGAGAACCCCTACAAACCCCGCTGTCGCATCAAGAAAGAGACAGCA 719
Qy 878 CCAAGTGGGCGAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGACTTCTGGG 937
Db 720 CCAAGTGGGCGAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGACTTCTGGG 779
Qy 938 AGGTGAGCTGGGCGCATCCCGCGCGCTGGAAGAGAGAGAGCGGTGACCGTGC 997
Db 780 AGGTGAGCTGGGCGCATCCCGCGCGCTGGAAGAGAGAGAGCGGTGACCGTGC 839
Qy 998 TGAAGTGGGCGACCGCTACTTTCAGCTGCGCGCTGAGCGAGCTTCCGCAAGTACACCG 1057
Db 840 TGAAGTGGGCGACCGCTACTTTCAGCTGCGCGCTGAGCGAGCTTCCGCAAGTACACCG 899
Qy 1058 CTTTCAACCATCCCGAGCATCAACAAAGAGACCGCGCATCCGCTACAGTACACAGTGC 1117
Db 900 CTTTCAACCATCCCGAGCATCAACAAAGAGACCGCGCATCCGCTACAGTACACAGTGC 959
Qy 1118 TGGCCCGAGGCTGGAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1177
Db 960 TGGCCCGAGGCTGGAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1019
Qy 1178 AGCCCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1237
Db 1020 AGCCCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1079
Qy 1238 GGAACCTGAGATCGGCGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1297
Db 1080 GGAACCTGAGATCGGCGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1139
Qy 1298 GCTGGGCGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1355
Db 1140 GCTGGGCGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1199

```

QY 1356 -----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGA 1411  
DB 1200 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGA 1259  
QY 1412 GCTGGAACGTAACGACATCCAGAACTGTGTGGGCAAGCTGAAGTGGGCCAGGACATCT 1471  
DB 1260 GCTGGAACGTAACGACATCCAGAACTGTGTGGGCAAGCTGAAGTGGGCCAGGACATCT 1319  
QY 1472 ACCCGGCGATCAAGTGTGGCGAGCTGTGCAAGCTGTGCGGGCGCCAAAGCCCTGACCG 1531  
DB 1320 AGCGCGCATCAAGTGTGAGAGCTGTGCAAGCTGTGCGGGCAGCCAGGCCCTGACCG 1379  
QY 1532 ACATGTCGCCCTCAACGAGGAGCCGAGCTGTGAGCTGTGCGGAGAAACGCGAGATCTGC 1591  
DB 1380 AGGTGATCCCCCTGACCGAGGAGCCGAGCTGTGAGCTGTGCGGAGAAACGCGAGATCTGA 1439  
QY 1592 GCGAGCCGCTGCAGCGGTGTACTACGACCCAGGAGACCTGTGTCGCCGAGATCCAGA 1651  
DB 1440 AGGAGCCGCTGCAGAGGTGTACTACGACCCAGGAGACCTGTGTCGCCGAGATCCAGA 1499  
QY 1652 AGCAGGCCACGACCACTGGACCTTACCAGATCTTACCAGGAGCCCTTCAAGAACTGAAGA 1711  
DB 1500 AGCAGGCCACGACCACTGGACCTTACCAGATCTTACCAGGAGCCCTTCAAGAACTGAAGA 1559  
QY 1712 CCGGCNAGTACGCAAGTGCACCGCCGACCCACACCAAGCAGTGAAGCAGCTGACCGAGG 1771  
DB 1560 CCGCAAGTACGCCCCGATGCGCGGCCACCAACAGCTGAAGCAGCTGACCGAGG 1619  
QY 1772 CCGTGCAGAACTGCCATGAGAGCATCGTGTACTGTGGGCAAGACCCCAAGTTCGCGC 1831  
DB 1620 CCGTGCAGAACTGAGCACCGAGAGCATCGTGTACTGTGGGCAAGATCCCCAAGTTCAGC 1679  
QY 1832 TGCCATCCAGAGAGACCTGGAGAGACCTGGTGGACCGACTACTGCGAGGCCACCTGGA 1891  
DB 1680 TGCCATCCAGAGAGACCTGGAGAGCCTGGTGGATGGAGTACTGGGAGGCCACCTGGA 1739  
QY 1892 TCCCGAGTGGAGTTCGTGAACACCCGCCCTGTGTGAAGCTGTGTACTGAGCTGGAGA 1951  
DB 1740 TCCCGAGTGGAGTTCGTGAACACCCGCCCTGTGTGAAGCTGTGTACTGAGCTGGAGA 1799  
QY 1952 AGGAGCCCATCATCGGCGCGAGACCTTCTACGTGTGAGCGCGCCGCCAAACGCGAGACCA 2011  
DB 1800 AGGAGCCCATCATCGGCGCGAGACCTTCTACGTGTGAGCGCGCCGCCAAACGCGAGACCA 1859  
QY 2012 AGATCGGCAAGCGCGCTGACCGACCGGGCGCGGAGAGATCGTGAAGCTGACCG 2071  
DB 1860 AGCTGGGCAAGCGCGCTGACCGACCGGGCGCGGAGAGATCGTGAAGCTGACCG 1919  
QY 2072 AGACCAACCAACAGAGACCGAGCTGCAGGCGCATCCAGCTGGCCCTGCAGGACAGCGGCA 2131  
DB 1920 ACACCAACCAACAGAGACCGAGCTGCAGGCGCATCCACCTGGCCCTGCAGGACAGCGGCC 1979  
QY 2132 GCGAGGTGAACATCGTGAACGACAGCCAGTACGCCCTGGGCATCATCAGGCCCGAGCCCG 2191  
DB 1980 TGGAGGTGAACATCGTGAACGACAGCCAGTACGCCCTGGGCATCATCAGGCCCGAGCCCG 2039  
QY 2192 ACAGAGCGAGAGCGAGCTGTGAGCCAGATCATCGAGCAGCTGATCAAGAGAGAGAGG 2251  
DB 2040 ACAGAGCGAGAGCGAGCTGTGAGCCAGATCATCGAGCAGCTGATCAAGAGAGAGAGG 2099  
QY 2252 TGTAACCTGAGTGGGTGCCGCCCAAGGGGATCGCGGCAACAGAGCAGATCGACCAAGC 2311  
DB 2100 TGTAACCTGAGTGGGTGCCGCCCAAGGGGATCGCGGCAACAGAGCAGTGGAGCAAGC 2159  
QY 2312 TGGTGAACAGGGGATCGCGAAGGTGTGTTCTGGAGCGGATCGATGCGCGGCTCGTGA 2371  
DB 2160 TGGTGAACAGGGGATCGCGAAGGTGTGTTCTGGAACGGATCGATGCGCGGCTCGTGA 2219  
QY 2372 TCTACAGTACATGACGACCTGTACGTGGGCGAGCGCGGCCCTAGGATCGATTAAAGC 2431  
DB 2220 TCTACAGTACATGACGACCTGTACGTGGGCGAGCGCGGCCCTAGGATCGATTAAAGC 2279

QY 2432 TTCGCGGGCTAGACACCGGTGAATTC 2457  
DB 2280 TTCGCGGGCTAGACACCGGTGAATTC 2305

RESULT 7  
US-09-552-950-2  
; Sequence 2, Application US/09552950  
; Patent No. 6541248  
; GENERAL INFORMATION:  
; APPLICANT: Oxford Biomedica (UK) Limited  
; TITLE OF INVENTION: Anti-Viral Vectors  
; FILE REFERENCE: 674524-2004  
; CURRENT APPLICATION NUMBER: US/09/552,950  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4307  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:gagpol-SYNgp - codon  
; OTHER INFORMATION: optimised gagpol sequence  
US-09-552-950-2

Query Match 66.2%; Score 1626.6; DB 4; Length 4307;  
Best Local Similarity 82.2%; Pred No. 1.4e-263;  
Matches 1951; Conservative 0; Mismatches 394; Indels 28; Gaps 6;

QY 12 CATGCGCGAGGCCATGAGCCAGG---CCACGAGCGCCAAATCTCTGATGCGAGCGGAGCAA 68  
DB 1086 CTTGGCTGAGGCCATGAGCCAGGTCACCAATCCCGTACCATCATGATGCGAGCGCGGCAA 1145  
QY 69 CTTCAAGGCCCCCAAGCGCATCATCAAGTGTTCACCTGCGCAGAGGAGGCGCACATCGC 128  
DB 1146 CTTTCGGAACCAACGCAAGATCGTCAAGTGTTCACCTGTGGCAAGAGGCGCACACAGC 1205  
QY 129 CCGCAACTCGCGCGCCCGCCGCAAGAAGGCTGTCTGGAAGTCCGGCAAGAGGCGGCAACA 188  
DB 1206 CCGCAACTCGAGGCGCCCTAGGAAAAGGGCTGTCTGGAATGCGGCAAGAGGCGCAACA 1265  
QY 189 GATGAAGACTGACCGAGCGCGAGCCCAATCTTCTTCGCGAGGACCTGGCCTTCCGCCA 248  
DB 1266 GATGAAGACTGACCGAGAGAGAGGCTAA-TTTTATAGGGAAGATCTGGCCTTCTTACA 1324  
QY 249 GGGCAAGGCCCGCGAGTTCCCGAGGAGCAGAAACCGCGCCCAACAGCCCGCCACGAGCGCGA 308  
DB 1325 AGGGAAGGCCAGGGAATTTTCTTCAAGCAGACCCCGCGCCACACAGCCCGCGCGA 1384  
QY 309 GCTGCAGGT-----GCGCGCGCAACCCCGCAGCGAGCGCGCGCGCGAGCGCGAGG 362  
DB 1385 GCTTCAGGTCTGGGGTGGGCAACAACATCTCCCTCCGAAGCAGGAGCGGACCGCGAGG 1444  
QY 363 CA-----CCCTGAATTCCTCCAGATCACCTCTGGCGAGCGCCCTCTGGTGAATCAA 416  
DB 1445 CAGGTGTCTTCAACTCTCCCTCAGGTACCGTCTTTGGCAGCGACCCCTGTGTCACATCAA 1504  
QY 417 GGTGGCGGCCAGATCAAGGAGGCCCTGTGTGACACCGCGCGCGAGCACACCGTGTGGA 476  
DB 1505 GATCGGGGGGAGCTCAAGAGGCTCTCTGTGACACCGCGAGCAGACACCGTGTGGA 1564  
QY 477 GGAGATGAGCTGCGCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGGCTTCA 536  
DB 1565 GGAGATGTCGTTGCCAGGCGCTGGAAGCCGAAGATGATCGGGGGAATCGGGGCTTCA 1624  
QY 537 CAAGTGTCCCGAGTACGACGATCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCAC 596  
DB 1625 CAAGGTGCGCGAGTATGACGATCTCTCAAAATCTGCGGCCACAGGCTATCGGTAC 1684  
QY 597 CGTGTGATCGGCGCCACCCCGTGAACATCATCGCCCGCAACATGCTGACCGAGTGG 656  
DB 1685 CGTGTGTGGGGCCCGACACCGCTCAACATCATCGGACGCAACCTTTGACGAGATCG 1744



QY 69 CTTCAAGGGCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGGCCACATCGC 128  
 Db 2253 CTTTCGGAACAAGCAAGATCTCAAGTCTTCAACTGCGGCAAGAGGGCCACACAGC 2312  
 QY 129 CCGCAACTGCGGCGCCCGCGCAAGAGAGGCTCTGGAAGTGGCGCAAGAGGGCCACCA 188  
 Db 2313 CCGCAACTGCGGCGCCCGCTAGGAAAGAGGCTCTGGAATGTGGAAGAGAGGACCA 2372  
 QY 189 GATGAAGGATGTGACCGAGCGCCAGAGGCAACTTCTTCGCGAGGACTGGCCCTTCCCCCA 248  
 Db 2373 AATGAAGATTTGACTGAGAGACAGGCTAA-TTTTATTGGAAGATCTGGCCCTTCCACA 2431  
 QY 249 GGGCAAGCGCGGAGTTCCTCCAGCGAGAGAAACCGCGCCCAACAGGCCCAACGAGCGCA 308  
 Db 2432 AGGGAAGCGCGAGGAAATTTCTTCAGAGCAGACAGAGGCAACAGGCCCAACAGAGAGA 2491  
 QY 309 GCTGACGCTGCGG-----CGACAAACCCCGCAGCGAGCGCGCGCGCGAGCGCCAGGG 362  
 Db 2492 GCTTCAGGTTTGGGAAGAGACAACTCCCTCTCAGAGCAGAGCGATAGACAGG 3551  
 QY 363 CA-----CCCTGAACTTCCCGCAGATCAACCTGTGTGGCAGCGCCCTGTGTAGCATCAA 416  
 Db 2552 AACTGTATCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCAATAAA 2611  
 QY 417 GGTGGCGGCGAGATCAAGAGGCGCTGTGGACACCGCGCGCGAGCAGACACCGTGTGGA 476  
 Db 2612 GATAGGGGGGAGCTCAAGAGGCTCTCTGGACACCGGAGCAGACGACCTGTCTGGA 2671  
 QY 477 GGAGATGAGCTGCGCGCAAGTGGAAAGCCCAAGATGATCGCGGCGATCGCGCGCTTCAT 536  
 Db 2672 GGAGATGCTGTTGCCAGCGCTGGAAGCCGGAAGATGATCGGGGGAATCGCGGCTTCAT 2731  
 QY 537 CAAGGTGGCGAGATCAAGACAGATCTGTATCGAGATCTGGGCGCAAGAGGCCATCGGCAC 596  
 Db 2732 CAAGGTGGCGAGATGACCAAGATCTCTCATCGAAATCTGGCGCGCAAGAGCTATCGGTAC 2791  
 QY 597 CGTGTGATCGGCGCCCAACCGCGTGAACATCATCGGCGCAACATGTGTGACCCAGCTGGG 656  
 Db 2792 CGTGTGTGGCGCCCAACCGCTCAACATCATCGAGCGCAACCTGTTGACGCGAGATCGG 2851  
 QY 657 CTGCACTGAACTTCCCATCAGCCCATCGAGACCGTGGCCGTGAGCTGAGCCGCG 716  
 Db 2852 TTGACGCTGAACTTCCCATTTAGCCCTATCGAGACGCTACCGGTGAGCTGAGCCGCG 2911  
 QY 717 CATGAGCGGCGCAAGGTGAGAGTGGCCCTTCACCGAGGAGAGATCAAGGCGCTGAC 776  
 Db 2912 GATGAGCGGCGCAAGGTCAAGCAATGGCCATTCACAGAGGAGAGATCAAGGCACTGGT 2971  
 QY 777 CGCCATCTGCGAGGAGATGGAAGAGGCGCAAGATCACCAAGATCGGCGCCCGAGAACCC 836  
 Db 2972 GGAGNTTGCACAGATGGAAGAGGAGGAATCTTCAAGATTTGGGCTGAGAACCC 3031  
 QY 837 CTACAAACCCCGTGTTCGCGATCAAGAGAGAGCAGACCAAGTGGCGCAAGCTGGT 896  
 Db 3032 GTACAAACCGCGGTGTTTCGCAATCAAGAGAGGAGCTCGACGAATGGCGCAAGCTGGT 3091  
 QY 897 GGACTTCGCGAGCTGAACAGCGCACCGAGCTTCTGGAGGTGAGCTGGGCACTCCC 956  
 Db 3092 GGACTTCGCGAGCTGAACAGCGCACCGAGCTTCTGGAGGTTTCACTGGGCACTCCC 3151  
 QY 957 CCACCCCGCGCTTGAAGAGAGAGAGCGGTGACCGTGTGAGAGTGGCGGAGCGCTA 1016  
 Db 3152 GCACCCCGCAGGGCTGAAGAGAGAAATCCGTGACCGTACTGAGATGTGGGTGATGCTA 3211  
 QY 1017 CTTGAGCTGGCCCTGAGCGAGCTTCGCAAGTACACCGCTTCACCATCCCGAGCAT 1076  
 Db 3212 CTTCTCCGTTCCCTTGGAGAGAACTTCAGGAAGTACACTGCCCTTCACATCCCTTCGAT 3271  
 QY 1077 CAACACAGAGACCCCGGCTATCCGCTACCAAGTACAACTGTGCGCCCAAGGCTGGAAGGG 1136  
 Db 3272 CAACACAGAGACACCGGGATTCGATATCAGTACAACTGTGCGCCCAAGGCTGGAAGG 3331  
 QY 1137 CAGCCCGCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGGCTTCCGCGCCCGCAA 1196

Db 3332 CTCTCCGCAATCTTCCAGAGTAGCATGACCAAAATCTTGAGGCTTTCGCAAAACGAA 3391  
 QY 1197 CCGCGAGATCGTATCTTACCA-----GGCCCCCTGTACGTGGCGAGCACCTGGAGAT 1250  
 Db 3392 CCGCGAATCGTCACTTATCAGTACATGATGACTTGTACGTGGCTCTGATCTAGAGAT 3451  
 QY 1251 CCGCGAGCACCGCGCAAGATCGAGGAGCTGCGCAAGACCTGCTGCTGGCTGGGCTTCAC 1310  
 Db 3452 AGGGCAGCACCGCGCAAGATCGAGGAGCTGCGCAGACCTGTGAGGTGGGAGCTGAC 3511  
 QY 1311 CACCCCGCAAGAGACCAAGAGGAGCGCCCTTCTGCCCCAT-----CGAGCTGCA 1364  
 Db 3512 CACACCGCAAGAGACCAAGAGGAGCTTCCCTTCTCTGGATGGTTTACGAGCTGCA 3571  
 QY 1365 CCGCGAAGTGGACCGGTGAGCCCTGAGCTGCGCGAGAGAGAGCTGAGACCGTGA 1424  
 Db 3572 CCCTGACAAATGGACCGGTGAGCCCTATCGTGTCTCCAGAGAAAGACAGCTGGA 3631  
 QY 1425 CGACATCCAGAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGCATCAA 1484  
 Db 3632 CGACATCAGAGCTGTGGGGAAGTTGAATGGGCCAGTCAGATTTACCCAGGATTA 3691  
 QY 1485 GGTGCGCAGCTGTGCAAGCTGTGCGCGCGCCCAAGCCCTGACCGACATCTGTCCTCT 1544  
 Db 3692 GGTGAGGAGCTGTGCAAACTCTCCGCGGAACCAAGGCACTCACAGAGGTGATCCCTCT 3751  
 QY 1545 GACGAGAGGCGGAGCTGGCGGAGAACCGCGAGATCTGCGCGAGCCCGTGA 1604  
 Db 3752 AACCGAGAGGCGGAGCTGGAATCGGCAAGAACCGAGAGATCTTAAGAGAGCCCGTGA 3811  
 QY 1605 CGGCGTGTACTACGACCCCAAGGACCTGGTGGCGAGATCCAGAGAGCGGCGCACGA 1664  
 Db 3812 CGGCGTGTACTATGACCCCTTCCAGGACCTGATCCCGAGATCCAGAGAGCGGCGCAAG 3871  
 QY 1665 CGAGTGGACCTACAGATCTACCGAGGCGCTTCAAGAACCTGAGACCGCGCAAGTACGC 1724  
 Db 3872 CCAGTGGACCTATCAGATTTACCGAGGCGCTTCAAGAACCTGAGAACCGCGCAAGTACGC 3931  
 QY 1725 CAAAGTGGCGCACCGCGCCACACCAAGAGAGCTGACCGAGCGCGCTGCAAGAGAT 1784  
 Db 3932 CGGATGAGGGGTGCCACACTAACGACGTCAAGCAGCTGACCGAGGCGCTGCAAGAGAT 3991  
 QY 1785 CGCCATGAGAGAGATCTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAA 1844  
 Db 3992 CACCACCGAAGATCTGATCTGGGGAAGACTCTTAAGTTCAAGCTGCCATCCAGAA 4051  
 QY 1845 GGAGACCTGGGAGACCTGTGTGACCGACTACTGGGAGGCGCACCTGGATCCCGAGTGGGA 1904  
 Db 4052 GGAACCTGGGAAAACCTGTGTGACAGAGTATTGGCAGGCGCACCTGGATTTCTGAGTGGGA 4111  
 QY 1905 GTTCTGTGAACACCCCGCCCTGCTGAAGCTGTGTGTACCGAGCTGGAGAGGAGCCCATCAT 1964  
 Db 4112 GTTCTGTCAACACCCCTCCCTGCTGTGAAGCTGTGTGTACCGAGCTGGAGAGGAGCCCATAGT 4171  
 QY 1965 CCGCGCGGAGACCTTCTTACGTGGACCGCGCCCAACCGCGAGACCAAGATCGGCAAGGC 2024  
 Db 4172 GGGCGCGGAAAACCTTCTACGTGGATGGGCGCGCTAACAGGGAGACTAAAGCTGGGCAAGGC 4231  
 QY 2025 CGGCTACGTGACCGAGCGGCGCGCAGAGAGATCTGAGCCTGACCGAGACCAACCAACCA 2084  
 Db 4232 CGGATACGTCACTAACCGGGGCGACAGAGGTTGTACCCCTCACTGACACCAACCAACCA 4291  
 QY 2085 GAAGACCGAGCTGACAGGCGCATCGAGTGGCCCTGAGAGACAGCGGAGCGAGGTGACAT 2144  
 Db 4292 GAAGACTGAGCTGACAGGCGCATTTACCTCGCTTTGAGGAGCTCGGGCGCTGGAGGTGAACT 4351  
 QY 2145 CGTGAACGACAGCGAGTACGCTTGGCATCTCAGGCGCCAGCGCCGACAGAGCGAGAG 2204  
 Db 4352 CGTGAACAGACTCTCAGTATGCTTGGCATCTTCAAGCCCGAGCGACAGAGTGAATC 4411  
 QY 2205 CGAGCTGCTGGAACACAGATCATCGAGCAGCTGATCAAGAGAGGAGAGGTGTACTGAGCTG 2264

Db 4412 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAGGAAAGGTTCTATCTGGCCTG 4471  
Qy 2265 GGTGCCCGCCCAACAGGGCATCGCGGCAACGAGCAGATCGAAGAGCTGGTGGCAAGG 2324  
Db 4472 GGTACCGCCCAACAGGCAATGGCGCAATGAGCAGGTGCGAAGAGCTGGTCTCGCTGG 4531  
Qy 2325 CATCCCAAGGTGCTGTTCTCGACGGCATCGA 2357  
Db 4532 CATCAGAAAGGTGCTATTCTCGATGGCATCGA 4564

RESULT 9  
US-09-872-733A-6  
; Sequence 6, Application US/09872733A  
; Patent No. 6856706  
; GENERAL INFORMATION:  
; APPLICANT: The Government of the United States of America, as  
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND  
; FILE OF INVENTION: SIV ENV GENES  
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL, SIV GAG & ENV  
; CURRENT APPLICATION NUMBER: US/09/872,733A  
; CURRENT FILING DATE: 2001-06-01  
; PRIOR APPLICATION NUMBER: PCT/US00/34985  
; BEST FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/173,036  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 8366  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA sequence  
; OTHER INFORMATION: of the construct pCMVgagpolBkan containing a CMV  
; OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin  
; OTHER INFORMATION: resistance gene  
US-09-872-733A-6

Query Match 62.6%; Score 1539.2; DB 4; Length 8366;  
Best Local Similarity 79.9%; Pred. No. 6.2e-249;  
Matches 1896; Conservative 0; Mismatches 448; Indels 28; Gaps 6;  
Qy 14 TGGCCGAGGCATGAGCGGCGACAGC---GCCACATCTTGATGCGAGCGAGCACT 70  
Db 1857 TGGCCGAGGCATGAGCGGCGACAGCAGTACGCGGACCAATATGATGCGAGAGGCACT 1916  
Qy 71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTCGGCAAGGAGGCGCCACATCGCCC 130  
Db 1917 TCCGGAACCGCGGAAGATCGTCAAGTGCTTCAATTGTGCAAGAGGCGCACACGCCA 1976  
Qy 131 GCAACTGCGCGCCCCCGCGAGAGGCTGCTGGAAGTGGCGCAAGGAGGCGCACCGA 190  
Db 1977 GGAAGTCTGCGGCGCCCCCGGAGAGGCTGTGGGAAATGTGGAAGAGGAGGACCCAAA 2036  
Qy 191 TGAAGGACTGACCGAGCGCGCAGCCCACTTCTTCCGCGAGGACCTGGCCCTTCCCCCAGG 250  
Db 2037 TGAAGATTGACTGAGAGACAGCTAA-TTTTITAGGAGATCTGGCCCTTCTCTACAAG 2095  
Qy 251 GCAAGGCGCGAGTTCCTCAGCAGAGAGACCGCGCAACAGCGCCACAGCGCGGAGC 310  
Db 2096 GGAAGGCGAGGAAATTTCTTCAGAGCAGACAGAGCCACAGCGCCACAGAGAGAGC 2155  
Qy 311 TGCAGGTGCGCG-----CGACAAACCCCGAGGCGCGCGCGCGAGCGCGCAGGGCA 364  
Db 2156 TTCAGTCTGGGTAGAGACAACTCTCCCTCAGAGCAGAGCGCGATAGACAGGAA 2215  
Qy 365 -----CCTGAATCTCCCGAGATCACTCTGTGGAGCGCGCCCTGGTGGAGCATCAAG 418  
Db 2216 CTGTATCTTTAACTTCCCTCAGATCACTTTTGGCAACGACCCCTCGTCAAGTAAAGGA 2275  
Qy 419 TGGCGGCGCAGATCAAGAGGCGCTGTGGACACCGGCGCGCAGCACCGCTGTGGAGG 478

Db 2276 TCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGCAGATGATACAGTATTAGAAG 2335  
Qy 479 AGATGAGCTTGCCTCCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCA 538  
Db 2336 AAATGAGTTTGCAGGAAGATGGAAACCAAAATGATAGGGGGGATTCGGGGCTTTCATCA 2395  
Qy 539 AGGTGGCGCCAGTACGACCCAGATCCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCACCG 598  
Db 2396 AGGTGAGGAGTACGACCCAGATCTCATAGAAATCTGTGACATAAAGCTATAGGTACAG 2455  
Qy 599 TGCTGATCGGCCCCCAACCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCT 658  
Db 2456 TATTAGTAGGACCTACACCTGTCAACATAATTGGAAAGAAATCTGTGTGACCCAGATCGGCT 2515  
Qy 659 GCACCTGAACTTCCCCCATCAGCCCATCAGAGCCCTGCGGCAAGCTGGAAGCTGAAGCCCGGCA 718  
Db 2516 GCACCTTGAATTTCCCATCAGCCCTATTGAGACGCTGCGCGTGAAGTGAAGCCCGGGA 2575  
Qy 719 TGGAGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCG 778  
Db 2576 TGGACGCGCCCAAGGTCAAGCAATGGCCATTGACGAAAGAGAGATCAAGGCCCTTAGTCG 2635  
Qy 779 CCATCTCGAGGAGATGCGAAGAGGCGCAAGATCACCAAGATCGGCCCGCGAAGCCCT 838  
Db 2636 AAATCTGTACAGATGCGAAGAGGAGGAGATCAGAAAGTCCGGCTTGAAGCCCT 2695  
Qy 839 ACAACACCCCGCTGTTCCCATCAAGAAAGACAGCACCAGTGGCGCGCGAGCTGGTG 898  
Db 2696 ACAACACTCCAGTCTTCGCAATCAAGAAAGAGCAGTACCAAGTGGAGAAAGCTGGTG 2755  
Qy 899 ACTTCCGCGAGCTGAACCAAGCGACCCAGGACTTCTGGGAGGTGCGAGCTGGGATCCCC 958  
Db 2756 ACTTCAGAGCTGTAACAAGAACTCAGGACTTCTGGGAGTTTCACTGGGATCCAC 2815  
Qy 959 ACCCGCGCGCTTGAAGAAAGAGCTGACCGTGTGAGCTGGAGCTGGGCGAGCCTTACT 1018  
Db 2816 ATCCCGCTGGGTTGAAGAAAGAGTCACTGACAGTGTGGTGTGGTGTGGTGTGGTGTGGT 2875  
Qy 1019 TCAGCTGCCCTCGAGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCCAGCATCA 1078  
Db 2876 TCTCGTTCCTTGGACAGGACTTCAAGAGTACATGCTTCCATACCATACCTAGCATCA 2935  
Qy 1079 ACAACGAGACCCCGGCTACCGCTACAGTACAAAGTGTGCTGCCCGAGGCTGGAGGGCA 1138  
Db 2936 ACAACGAGACACCGGCTACCGCTACAGTACAAAGTGTGCTGCCACAGGAGTGGAGGGAT 2995  
Qy 1139 GCGCGAGCATCTTCCAGAGCAGATGACCAAGTCTTGGAGCCCTTCCGGCGCGCGCAAC 1198  
Db 2996 CACCGAGCATCTTTCAAAGCAGCATGACCAAGTCTTGGAGCCCTTCCGCAAGCAAAAC 3055  
Qy 1199 CCGAGATCGTATCTTACCA-----GGCCCCCTGTAGCTGGGCGAGCGACTGGAGATCG 1252  
Db 3056 CAGACATCGTATCTTACGTATACATGAGCAGCTTACGTAGGAAGTGAACCTGGAGATCG 3115  
Qy 1253 GCGAGACCGCGCAGATCGAGAGCTGCGAAGCCTTCTGCTGGCTGGGCTTTCACCA 1312  
Db 3116 GCGACACAGGACCAAGATCGAGAGCTGAGACAGCTCTTTGAGGTGGGAGCTGACCA 3175  
Qy 1313 CCCCCGACAAGAGCACCAAGAGGAGCCCGCTTCTGCCCAT-----CGAGTGGCAC 1366  
Db 3176 CACGACAGAGAGCACCAAGAGGAGCTTCTTCTTCTGTGGATGGGCTTACGACTGATC 3235  
Qy 1367 CCGAAGAGTGGACCTGCGAGCCCATCGAGCTGCGCGAGAGAGCTGGACCGTGAACG 1426  
Db 3236 CTGCAAGTGGACAGTGGAGCCCATCGTGTGCTGAGAGGAGCAGCTGGAGTGTGAAG 3295  
Qy 1427 ACATCCAGAGCTGGTGGGAGCTGAACTGGGCGAGCGCAGATCTTACCCCGGCTTCAAG 1486  
Db 3296 ACATCAGAGCTCGTGGCAAGTGTGAACTGGGCAAGCCAGATCTTACCGAGCATCAAG 3355  
Qy 1487 TGGCGCAGCTGTGCAAGCTGTGCGCGCGCCAAAGGCCCTGACCGACATCGTGGCCCTGA 1546  
Db 3356 TTAGGCGCTGTGCAAGCTGTCTTCGAGGAAACCAAGGCACTGACAGAGTGTATCCCATGA 3415

QY 1547 CCGAGAGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCGCGAGCCGCTGCAGG 1606  
Db 3416 CAGAGGAAGCAGAGTAGAACTGGCAGAGAACCGAGAGATCTCTGAAGAGCCAGTACATG 3475  
QY 1607 GCGTGTACTACGACCCAGCAAGGAGCTGTGGCCGAGATCCAGAGGAGGAGCCAGCACC 1666  
Db 3476 GAGTGTACTACGACCCAGCAAGGAGCTGTGGCCGAGATCCAGAGGAGGAGCCAGCACC 3535  
QY 1667 AGTGGACCTACGAGATCTACGAGGAGCCCTTCAAGAACCTGAAGCCGCAAGTACGCCA 1726  
Db 3536 AATGGACCTACCAAAATCTACGAGGAGCCCTTCAAGAACCTGAAGCCGCAAGTACGCCA 3595  
QY 1727 AGATGGCAGCCGCGCCACACAGAGCTGAAGCAGCTGACCGAGGCGCTGCGAGAGATCG 1786  
Db 3596 GATGAGGGGTGCCACACACAGATGTGAGCAGCTGACAGAGGCGAGTGCAGAGATCA 3655  
QY 1787 CCATGAGAGCATCTGTATCTGGGGCAAGACCCCAAGTTCGCGTCCGCCATCCAGAGG 1846  
Db 3656 CCACAGAGAGCATCTGTATCTGGGGCAAGACTCCCAAGTTCGCGTCCGCCATCCAGAGG 3715  
QY 1847 AGACCTGGGAGACTGTGTGAGCAGCTACTTGGCAGGCGACCTGGATCCCGAGTGGAGT 1906  
Db 3716 AGACATGGGAGACATGTGTGAGCAGCTACTTGGCAGGCGACCTGGATCCCGAGTGGAGT 3775  
QY 1907 TCGTGAACACCCCGCCCTCGTGAAGCTGTGTGTAACAGCTGGAGAGGAGCCCATCATCG 1966  
Db 3776 TCGTGAACACCCCGCCCTCGTGAAGCTGTGTGTAACAGCTGGAGAGGAGCCCATCATCG 3835  
QY 1967 GCGCCGAGACTTCTAGTGGAGCGCGCGCCCAACCGGAGACCAAGATCGGCAAGCGCG 2026  
Db 3836 GAGCAGAGACTTCTAGTGGATGGGCGAGCCCAACCGGAGACCAAGTGGCGGAGCGAG 3895  
QY 2027 GCTACGTGACGACCGGCGCGCGCAGAGATCTGTAGCTGTGACCGGAGACCAACAGCA 2086  
Db 3896 GCTACGTGACGACCGGAGGAGCAGAGAGTGTGTGACCGCTGTGACACCAACAGCA 3955  
QY 2087 AGACCGAGCTGCAGGCGCATCCAGCTGCGCTGAGGAGCAGCGGAGGAGTGAACATCG 2146  
Db 3956 AGACTGAGCTGCAAGCCATCTACTAGCTCTGCAAGCAGCGAGCTGGAAGTGAACATCG 4015  
QY 2147 TGACCGCAGCAGCTAGTGGCCCTGGGATCATCCAGGCGCCAGCCGCGAGAGCGAGCGG 2206  
Db 4016 TGACAGACTCAGACTAGCTGGGATCATCCAGGAGCAGCAACAGGAGTCCAGTCCAG 4075  
QY 2207 AGCTGGTGAACAGATCATCGAGCAGCTGTATCAAGAGGAGAGGTTGACTGAGCTGGG 2266  
Db 4076 AGCTGGTGAACAGATCATCGAGCAGCTGTATCAAGAGGAGAGGTTGACTGAGCTGGG 4135  
QY 2267 TGCCCGCCACAGGGGAGTGGGGCAACGAGCAGATCGACAGCTGGTGAAGGAGGCA 2326  
Db 4136 TACCAGCAGCAAGAGGAATGGAGGAATGAACAGTAGATTAATAGTCAGTGGTGGGA 4195  
QY 2327 TCCGCAAGGTGCTGTCTCTGACCGCATCGAT 2358  
Db 4196 TCCGCAAGGTGCTGTCTCTGACCGCATCGAT 4227

RESULT 10  
US-09-872-733A-1  
; Sequence 1, Application US/09872733A  
; Patent No. 6656706  
; GENERAL INFORMATION:  
; APPLICANT: The Government of the United States of America, as  
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND  
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL, SIV GAG & ENV  
; CURRENT FILING DATE: 2001-06-01  
; PRIOR APPLICATION NUMBER: PCT/US00/34985  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/173,036  
; PRIOR FILING DATE: 1999-12-23

; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4338  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Mutated Human  
; OTHER INFORMATION: Immunodeficiency Virus - 1 Gag/Pol gene  
US-09-872-733A-1  
  
Query Match 61.28; Score 1503.8; DB 4; Length 4338;  
Best Local Similarity 79.84; Pred. No. 4,7e-243;  
Matches 1896; Conservative 0; Mismatches 447; Indels 34; Gaps 9;  
  
QY 14 TGGCCGAGGCGCATGAGCCAGGCCACCAGC---GCCAATCATCTCTGATGACCGCAGCAACT 70  
Db 1085 TGGCCGAGGCGATGAGCCAGGTGAGAACTCGGCGACCATATGATGACAGAGGCAACT 1144  
QY 71 TCAGGGGCGCCAGACCGCATCATCAAGTGTCTCAATCGCGCAAGGAGGCGCATCGCCC 130  
Db 1145 TCCGGAACACAGCGGAAGATCGTCAAGTGTCTCAATTTGTGCAAGAGGCGCACCGCCA 1204  
QY 131 GCAATCGCGCGCCCGCCCGCAAGAGGGCTGTGGAAAGTGGCGCAAGAGGGCGCCACAGA 190  
Db 1205 GGAATCGCGCGCCCGCCCGCAAGAGGGCTGTGGAAATGTGGAAGAGGACACCAAA 1264  
QY 191 TGAAGGACTGACCGAGCGCCAGCCAACTTCTTCCCGAGGACCTGGCTTCCCGCAGG 250  
Db 1265 TGAAGATTTGACTGAGAGACAGGCTAA-TTTTATTAGGGAAGATCTGGCTTCTCTAAG 1323  
QY 251 GCAGGCGCGGAGTTCCTCCAGCGAGCAGAAACCGCGCAACAGCCCGCCAGCGCGGAGC 310  
Db 1324 GGAAGCGCGAGGAATTTCTTCAGGAGCAGACAGGCGCAACAGCCCGCCAGAGAGAGC 1383  
QY 311 TGCAGGTGCGCG-----CGACAAACCGCGAGCGAGCGCGCGCGCGCGCGCGCGCA 364  
Db 1384 TTCAGGTCTGGGGTAGAGACAACTCCCGCTCAGAAGCAGGAGCGGATAGACAGGAA 1443  
QY 365 -----CCCTGAACTTCCCGCAGATCACCTGTGCGAGCGCCCGCTGTGAGCATCAAG 418  
Db 1444 CTGTATCTCTTAACTTCTCTCAGATCTCTTTGGCAAGCCCTCTGTCTACAGTAGGA 1503  
QY 419 TGGCGGCGCAGATCAAGAGGCGCTGTGGAACACCGCGCGCGAGCAGCACCGTGTGAGG 478  
Db 1504 TCGGGGGCACTCAAGAGGCGCTGTGATACAGGAGCAGATGATACAGTATTAGAG 1563  
QY 479 AGATGAGCTTCCCGGCAAGTGGAAAGCCCAAGATGATCGCGGCGATCGCGGCTTCACTA 538  
Db 1564 AAATGAGTTTCCAGGAAGATGGAACCAAAATGATAGGGGGGATCGGGGGCTTCACTA 1623  
QY 539 AGGTGCGCAGTACGACGAGATCTGTGAGATCTGCGCAAGAGGCCATCGGCACCG 598  
Db 1624 AGGTGAGGAGTACGACGAGATCTCATGAAATCTGTGAGCATTAAGCTATAGGTACAG 1683  
QY 599 TGCTGATCGGCGCCACC-----CCCGTGAACATCATCGCGCCCAACATGCTGACCGAGCTG 654  
Db 1684 TATTAGTAGGACCTACCTACACCTGTCAACATAATTGGAAGAAATCTGTTCACCCAGATC 1743  
QY 655 GGCTGACCTGAACTTCCCGCATGAGCCCATCGAGACCGTGGCGGTGAAGTGAAGGCC 714  
Db 1744 GGCTGACCTGAACTTCCCGCATGAGCCCATGAGACCGTGGCGGTGAAGTGAAGGCC 1803  
QY 715 GGCAATGAGCGCGCCCAAGGTGAAGCAGTGGCGCTGACCGGAGGAGAGATCAAGGCCCTG 774  
Db 1804 GGGATGAGCGCGCCCAAGGTCAAGCAATGGCCATTTGACAAAGAGAGATCAAGGCCCTTA 1863  
QY 775 ACCGCGATCTGGAGGAGATGAGAGGAGGAGGAGATCAACAGATCGGCGCGGAGAAC 834  
Db 1864 GTCGAAATCTGTACAGAGATGAGAGGAGGAGGAGATCAGCAAGATCGGCGCTGAGAAC 1923  
QY 835 CCTTACACACCGCGCTGTTCGCCATCAAGAGAGGAGCAGCAACCAAGTGGCGCAGCTG 894





Db 1401 AAGACTGTACTGAGAGGAGGTAA-TTTTTTAGGAAAAATTTGGCTTCCACAGGGA 1459  
Qy 254 AGCCCGCGAGTTCCCGAGAGCAGAAACCGCGCCAAACAGCCCAACAGCGCGAGTGC 313  
Db 1460 AGCCAGGGAATTTCTTTCAGAAACAGCAGAGCCAAACAGCCCAACAGCAGAGAGCTTC 1519  
Qy 314 AGGTGGCGGCGACAAACCCCGCAGAGCGCGCGCGCGAGCGCCAGAGGCAACCTTGAAC 373  
Db 1520 AGGTTCGAGAGACAACCCCGCTTCGAGCAGAGAGTTCGAAACAGAGGAAGCTTAACT 1579  
Qy 374 TCCCCCAGATCACCTGTGTCAGCGCCCTCTGTGAGCATCAAGGTGGCGCGCAGATCA 433  
Db 1580 TCCTCAATCACTCTTTGGCAGCAGCCCTTCTCAATAAAGGTAGGGGCGCAATAA 1639  
Qy 434 AGAGGCTCTGTGACAGCGCGCGCAGCAGCAGCAGCTGTGAGGAGAGTGAAGCTCCCG 493  
Db 1640 AGAGGCTCTTTAGACACGGGAGCAGGTGATAGATTAGAAATAAATTTGCCAG 1699  
Qy 494 GCAAGTGAAGCCCAAGATGATCGGGGATCGCGGCTTCATCAAGGTGCGCGAGTACG 553  
Db 1700 GCAATGGAACCAAAATGATAGGAGAAATGGAGGCTTATCGAAGTAAGACAATATG 1759  
Qy 554 ACCAGATCTGTGAGATCTGGGCAAGAGGCCATCGCACCGTGTCTGATCGGCCCA 613  
Db 1760 ATCAAAATACCTATGGAAATTTGGAAAAAGGCTATAGGTACAGTATTTAGTAGGACCTA 1819  
Qy 614 CCCCCTGAACATCATCGCGCGCAACATGTGACCCAGCTGGGCTGCACCTGAACCTCC 673  
Db 1820 CACTGTCAACATAATTTGGAAGAAATATGTGTACTCAGCTTGGATGCACATAAATTTTC 1879  
Qy 674 CCATCAGCCCATCGAGACCGTGTGAGAGCTGAGAGCGCGGATGGAGCGGCCCAAG 733  
Db 1880 CAATTAGTCTTATGAACTGTACAGTAATAATTAAGCCAGGAATGGATGGCCCAAGG 1939  
Qy 734 TGAAGCAGTGGCCCTCAGCAGAGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793  
Db 1940 TTAACAATGCCATTTGACAGAGAGAAATAAAGCTTTAAACAGCAATTTGTGAAGAA 1999  
Qy 794 TGGAGAGAGGCGCAAGATCACCAAGATCGCGCCCGAGAACCCCTACAAACCCCGTGT 853  
Db 2000 TGGAGAGGAGGAAAAATTTACAAAAATTTGGGCTGAAAAATCCATATAACACTCCAGTAT 2059  
Qy 854 TCGCATCAAGAGAGCAGACCAAGTGGCGCAAGCTGGTGAGCTTCCGCGAGCTGA 913  
Db 2060 TTGCCATATAAAGAGAGACAGTACTAAGTGGCGAAAAATTTAGTAGATTTTCAAGGAACTCA 2119  
Qy 914 ACAAGCGCACCCAGACTTTTGGAGGTGACGTGGGCTATCCCCCAACCCCGCGGCTGA 973  
Db 2120 ATAAAGAACTCAAGACTTTTGGAAAGTTCAATTAGAATACCAACCCAGAGGGTTAA 2179  
Qy 974 AGAAGAGAGAGCGTGACCGTGTGACGTGGCGGACGCTACTTTCAGCGTGCCTGG 1033  
Db 2180 AAAAGAAAAATCAGTGACAGTACTGGATGTGGGGATGCATATTTTTCAGTTCTTTAG 2239  
Qy 1034 ACAGGACTTCCGCAAGTACACCGCTTCACCATCCCGAGCATCAACAGAGAGCCCG 1093  
Db 2240 ATGAAAGCTTCAGGAAATATATCTGATTCACCATCTAGTACAAACAAATGAAACACAG 2299  
Qy 1094 GCATCCGCTACAGTACAACTGTGTCCTCCAGGCTGGAAGGCGAGCCCAAGCATCTTCC 1153  
Db 2300 GGATTAGATATCAATATAATTTGTCTTCCAGGAGTGAAGAGATCACCAGCAATATTC 2359  
Qy 1154 AGAGCAGCATGACCAAGATCTTGAGCCCTTCCGCGCCGCAACCCCGAGATCTGTATCT 1213  
Db 2360 AGAGTAGCATGACAAAAATCTTTAGAGCCCTTCAGGGCAGAAAAATCCAGACATATCTCT 2419  
Qy 1214 ACCA-----GGCCCCCTGTACGTGGGCGAGCAGCTGGAGATGGCCAGCACCGGCCA 1267  
Db 2420 ATCAATATATGATGACCTGTATGTAGGATCTGACTTAGAATAGGGCAACATAGAGCAA 2479  
Qy 1268 AGATCGAGAGCTGCGCAAGCACTGTCTGGCTGGGGCTTCCACACCCCGCAGAGAGC 1327  
Db 2480 AATAGAAAGATTAGAGAAACATCTTATAAAGTGGGGATTTTACCACACCAAGAGAAAC 2539

Qy 1328 ACCAAGAGGAGCCCCCTTCTCTGCCAT-----CGAGCTGCACCCCGCAACAGTGGACCG 1381  
Db 2540 ATCAGAAAGAACCCCTATTTCTTTGGATGGGTATGAACCTCCATCTGACAAATGGACAG 2599  
Qy 1382 TGCAGCCCATGAGCTCCCGAGAGAGAGAGCTGAGCGTGAACGACATCCAGAGCTGG 1441  
Db 2600 TACAGCTATACAGCTGCGCAAAAAAGATAGCTGGACTGTTAATGATATACAGAGTTAG 2659  
Qy 1442 TGGCAGAGCTGAACCTGGCGCAGCCAGATCTACCCCGGCATCAAGGTGCGCGAGCTGTGCA 1501  
Db 2660 TGGAAATAATTAACCTGGGCAAGTCAGATTTACGAGGGATTAAGTAAGCAACTTTGTA 2719  
Qy 1502 AGCTGCTGCGCGCGCCAAAGCCCTGACCGACATCTGTGCCCCCTGACCGAGAGGCCAGC 1561  
Db 2720 AACTCCTTAGGGAGCCAAAGCACTAACAGACATAGTACCATACTAACTGAAGAAGCAGAA 2779  
Qy 1562 TGGAGCTGGCGCGAAGCCGGAGATCTCGCGAGCGCCCTGCGCGAGCTGTACTACGACC 1621  
Db 2780 TAGAATTGGCAGAGAAACAGGAATTTTAAAGAACCAAGTACATGCGGTATTTATGACC 2839  
Qy 1622 CCAGCAAGGACCTGTGTGGCCGAGATCCAGAGCAGGSCCAGCAGCTGAGACCTACAGAA 1681  
Db 2840 CATCAAAAGACTTGTATAGCTGAATAACAGAAACAAAGGCGATGACCAATGGACATATCAA 2899  
Qy 1682 TCTACAGAGCCCTTCAAGAACCTGAAGACCGCGCAAGTACGCCAAGATGCGCACCGCCC 1741  
Db 2900 TTTACAGGAAACCAATTCAAAAATCTGAATAACAGGAAAGTATGCAAAATTAAGGACAGCCC 2959  
Qy 1742 ACACCAACGACGTGAAGCAGCTGACCCAGCGCTGCGAAGAGATCGCATGGAGAGCATG 1801  
Db 2960 ACATTAATGATGAACAGTTAACAGAGGCGAGTGAACAAATAGCCCTGGAGAGCATAG 3019  
Qy 1802 TGATCTGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGAGAGAGCCTGGGAGACCT 1861  
Db 3020 TAATATGGGAAAGATTTCTTAATTTAGACTACCCATCCAAAAAGAAACATGGGAAACAT 3079  
Qy 1862 GGTGACCGACTACTTGGCAGCGCCACTGATCCCGAGTGGGAGTTCTGTGAACCCCCC 1921  
Db 3080 GGTGACAGACTATTTGGCAAGCCACTGATTTCTGTAGTGGAGTTTCTTAATACCCCTC 3139  
Qy 1922 CCTGCTGAAGCTGTGTTACAGCTGGAAGAGGCCCATCATCGGCGCCGAGACCTTCT 1981  
Db 3140 TCTTAGTAAAAATTTAGTACAGCTGGAGAAAGAACCCATAGTAGGAGCAGAAACCTTCT 3199  
Qy 1982 AGTGAAGCGCCCGCCAGCAGACCAAGATCGGCAAGCGCGCTACGTTGACCGGACC 2041  
Db 3200 ATGTAGATGGAGCGCCCAATAGGGAATTAATTTAGGAAAGCAGGGTATTTACTGACA 3259  
Qy 2042 GGGCGCGCAGAGATCTGTAGCCTGACCGAGACCAACCAACAGAGAGCCGAGCTGCAAG 2101  
Db 3260 GAGGAAGGCAAAAAATTTGTACTCTAACTGAACAAACAAATCAGAGAGCTGAATTAACA 3319  
Qy 2102 CCATCCAGCTGGCCCTGCGAGGACAGCGCAGCGAGGTGAACATCTGTGACCGAGCCAGT 2161  
Db 3320 CAATTTTACCTAGCTTTGCAAGATTTCAGGATCAGAAAGTAAACATAGTAACTGACTCAGCT 3379  
Qy 2162 AGCCCTGGGCTATCCAGGCCCGCAGCCGCAAGAGCGAGAGCGAGCTGTGTGAACCCAGA 2221  
Db 3380 ATGCGTTAGGAATCATTTCAAGCACAATCCAGATTAAGTGAATCAGAGTTAGTCAACCAA 3439  
Qy 2222 TCATCAGCAGCTGTATCAAGAGGAAAGGTGTACTGTAGCTGGGTGCGCCGCCCAAGG 2281  
Db 3440 TAATAGAACAAATTAATAAGAAAGGAAAGGTTACTGTCTATGGGTACCGACACATAAG 3499  
Qy 2282 GCATCGCGCGCAGCAGAGATCGCAGCTGCTGAGCAAGGGCATCCGCAAGGTGCTGT 2341  
Db 3500 GAATTCGAGGTATGAAACAGGTAGATAAATTTAGTAAGCAAGGAATCAGGAAGTGTCTGT 3559  
Qy 2342 TCCTGACCGGCATCGA 2357  
Db 3560 TTCTAGTGAATAGA 3575

RESULT 12  
US-09-184-418C-9  
; Sequence 9, Application US/09184418C  
; Patent No. 6492110  
; GENERAL INFORMATION:  
; APPLICANT: Hahn, Beatrice  
; APPLICANT: Gao, Feng  
; APPLICANT: Shaw, George  
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN  
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1  
; FILE REFERENCE: D6287  
; CURRENT APPLICATION NUMBER: US/09/184,418C  
; CURRENT FILING DATE: 1999-11-02  
; NUMBER OF SEQ ID NOS: 112  
; SEQ ID NO 9  
; LENGTH: 8972  
; TYPE: DNA  
; ORGANISM: Human immunodeficiency virus type 1  
; FEATURE:  
; OTHER INFORMATION: isolate=962W/51.3; 137.1632:gag; 1419.4435:pol;  
; OTHER INFORMATION: 4380.4958:vif; 4898.5188:vpr; 5169.7814:tat;  
; OTHER INFORMATION: 5308.7928:rev; 5407.5667:vpu; 5585.8128:env;  
; OTHER INFORMATION: 8130.8753:nef  
US-09-184-418C-9

Query Match 46.7%; Score 1147.4; DB 4; Length 8972;  
Best Local Similarity 69.3%; Pred. No. 1.8e-183;  
Matches 1647; Conservative 0; Mismatches 696; Indels 32; Gaps 5;

QY	14	TGGCGGAGCCATGAGCAGCGACCGACACGCGCCACATCTTGATGCGAGCGCAGCAACTCA	73
Db	1214	TGGCTGAAGCAATGAGCCAAGTAACCAATACAAACATAATGATGCGAAAGCAATTTTA	1273
QY	74	AGGCGCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGCGCACATCGCCGCA	133
Db	1274	AAGCCCTAAAGAAATGTTAAATGTTTCAACTGTGCGAGGAAGGCATATAGCCAGGA	1333
QY	134	ACTCCGCGCCCCCGCAGAGAGGCTGCTGGAAGTCCGCGACAGGAGGGCCACCATGTA	193
Db	1334	ATTCCAGGGCTCTCTGGGAAAAAAGGCTGTTGGAAATGTGAAAGGAGGACACCAATGA	1393
QY	194	AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCCTCCCCAGGGCA	253
Db	1394	AAGACTGTACTGAGAGACAGCTAA-TTTTITAGGGAATTTTGGCTTCCCAAGAGGG	1452
QY	254	AGGCCCGGAGTTCCCGACGACAG-----AACCGGCGCCACAGCC	295
Db	1453	AGGCCGGGGAATCTTCTTCAGAACAGACGAGCCACAGCCCCACCAGTCTCCACAGCC	1512
QY	296	CCACGAGCGCGAGCTGCAGTGCAGCGCGCACAACCCCGCAGCGAGCGCGCGCCGAGC	355
Db	1513	CCACGACGAGAGCTTCAGGAGAGACACCCCTGCCCGAGGAGGAGCAGAGAA	1572
QY	356	GCCAGGACCCCTGAATTCCTCCCGAGATCACCTGTGGCAGCGCCCTCGTGTGAGATCA	415
Db	1573	GACAAGGAACCTTTAACTGCGCTCAAAATCACTTTTGGCAGCGACCCCTGTCTCAATAA	1632
QY	416	AGGTGGGCGGCAGATCAAGGAGGCGCTGTGGACACCGCGCGCAGACACCGTGTCTGG	475
Db	1633	AAGTAGGGGTCAGATAAGAGAGGCTCTCTTTGGATACAGGAGCAGATGATACAGTATTAG	1692
QY	476	AGGAGATGAGCCTGCCCGCAAGTGGAAAGCCAAAGATGATCGCGCGCATCGCGCGCTTCA	535
Db	1693	AAGAAATAAATTTGCCAGGAAAAATGGAAACCAAAATCATAGGAGGAATTTGAGGTTTA	1752
QY	536	TCAAGGTGCGCAGTACACACAGATCTTGATCGAGATCTCGGCGAAGAGGCCATCGCA	595
Db	1753	TCAAGTAGACAGATATGATCAATACTTTAGAAATTTTGGAAAAAAGGCTATAGGTA	1812
QY	596	CGTGCTGATCGGCCCCACCCCGTGAAATCATCATCGGCCGCAACATGTCGACCCAGCTGG	655
Db	1813	CAGTATTAGTAGCACTACACTGTCAACATAATTTGGAGAAATATGTTGACCCAGCTTG	1872

1724 CCAAGATCGCACCGCCACCAACCAACGACGCTGAACGAGCTGACCGAGCCGCTGCAGAGA 1783  
2952 CAAAAATGAGGACTGCCACACACTAATGATGTAACAGAGGCTGCAAAAAA 3011  
1784 TCSCCATGGAGAGCATGCTGATCTGGGCAAGACCCCAAGTTCCGCTGCCATCCAGA 1843  
3012 TAGCCATGGAAGAGCATAGTAATATGGGGAAGATTCTTAATTTAGGCTACCCATTCAA 3071  
1844 AGGAGACCTGGAGACCTGCTGGACCGACTACTGCGCAGGCCACCTGGATCCCGAGTGGG 1903  
3072 AAGAAACATCGGAGACATGGTGGACAGACTATTGGCAAGCCACCTGGATTCTCTGAGTGG 3131  
1904 AGTTCTGTGAACACCCCCCTGCTGTGAAGCTGTGTGTTACAGCTGAGAGAGGCCATCA 1963  
3132 AGTTGTGTTAACTCTCCCCCTCTAGTAAATTTATGGTACCACTGGAGAAAGAACCCATAG 3191  
1964 TCGGCGCCGAGACCTTCTACGTGACGCGCGCCCAACCGGAGACCAAGATCGGCAAGG 2023  
3192 CAGNAGCAGAACTTACTATTAGATGAGCAGCAATAGGGAACCTTAATAGGAAG 3251  
2024 CCGGCTACGTGACCGAACCGCGGCGCGGAGAGATCGTGAAGCTGACCGAGACCAACACC 2083  
3252 CAGGCTATGTTACTGACAGAGGAAGGCAAAAAATTTGTTACTTAACCTGAACCAACAATC 3311  
2084 AGAAGACCGAGCTGCGAGCCATCCAGCTGGCCCTGCAGACAGCGGCGAGGAGGTGAACA 2143  
3312 AAAAGACTGAATTACAGCAATTCAGTTAGCTTTGAGGANTTGAAGTCAAGATGAACA 3371  
2144 TCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCAGCCCGACAGAGCGAGA 2203  
3372 TAGTAACAGACTCACAGTATGCATTTAGGAATCATCCAGACACACACAGATGAAGTGAAT 3431  
2204 GCGAGCTGTTGACAGAGATCATCGAGCAGCTGTATCAGAGGAGAGAGTGTACCTGAGCT 2263  
3432 CAGAATTAGTCAATCAATAATAGAACAGTTGATATAAAGGAAGGTTTACCTGTGCAT 3491  
2264 GGGTCCCGCCGACAGGCGCATCGCGGCAACGAGCAGATCGACAAGCTGTTGAGCAAGG 2323  
3492 GGGTACACGACACACAAAGGAATTGAGGAATGAACAGTAGATAGATAAATTGTTAAGTAGT 3551  
2324 GCATCCGCAAGGTGCTGTTCTCTGAGCGGATCGAT 2358  
3552 GAATCAGGAAGTGTGTTCTTAGATGGAATAGAT 3586

RESULT 13  
US-09-184-418C-11  
; Sequence 11, Application US/09184418C  
; Patent No. 6492110  
; GENERAL INFORMATION:  
; APPLICANT: Hahn, Beatrice  
; APPLICANT: Gao, Feng  
; APPLICANT: Shaw, George  
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN  
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1  
; FILE REFERENCE: D6287  
; CURRENT APPLICATION NUMBER: US/09/184,418C  
; CURRENT FILING DATE: 1999-11-02  
; NUMBER OF SEQ ID NOS: 112  
; SEQ ID NO 11  
; LENGTH: 8959  
; TYPE: DNA  
; ORGANISM: Human immunodeficiency virus type 1  
; FEATURE:  
; OTHER INFORMATION: isolate=94IN476.104; 138.1613: "gag";  
; OTHER INFORMATION: "pol"; 4361.4939: "vif"; 4879.5169: "vpr";  
; OTHER INFORMATION: 5150.7782: "tat"; 5289.7939: "rev"; 5378.5638: "vpu";  
; OTHER INFORMATION: 5556.8129: "env"; 8131.8754: "nef"  
US-09-184-418C-11

Query Match 46.4%; Score 1140.2; DB 4; Length 8959;  
Best Local Similarity 69.2%; Pred. No. 2.9e-182;  
Matches 1632; Conservative 0; Mismatches 708; Indels 17; Gaps 5;

QY 14 TGGCCGAGGCCATGAGCCAGGCGCACCGGCGCAACATCCTGATGAGCGCGAGCAACTTCA 73  
DB 1216 TGGCTGAGGCAATGAGCCAAATCACATAG---TAAATTAATGATGAGAGAGGCAATTTTA 1272  
QY 74 AGGSCCCCAAGCGCATCATCAAGTCTTCACTCGGCAAGAGGCGCCACATCGGCCGCA 133  
DB 1273 AAGGCCCTTAAAGAAATTTGTTAAATGCTTCACTGTGGCAAGGAGGCGCATAGCCAGAA 1332  
QY 134 ACTGCGCGCCCGCCCGCAAGAGGCGTCTGGAAGTGGCGCAAGGAGGCGCCACAGATGA 193  
DB 1333 ATTGCAAGGCCCTTAGAAAAAGAGGCTGTGGAAATGTGGCAAGAGGACACCAATGA 1392  
QY 194 AGGACTGCAACCGAGCGCGAGCCAACTTCTCCGCGAGGACCTGGCCCTTCCCGCAGGGCA 253  
DB 1393 AAGACTGTACTGAGAGGCGAGGCTAA--TTTTTTAGGGAATTTTGGCCTTCCCAAGGGG 1451  
QY 254 AGSCCCCGAGTTCCCGCAGCGAGAGAACCGCGCCCAACAGCCCGCCACCGCCCGAGCTGC 313  
DB 1452 AGSCCAGGGAATTCCTTCAAAAGAGGCGAGAGCCCAACAGCCCGCCACGAGAGAGCTTC 1511  
QY 314 AGGTGCGCGCGACAAACCCCGCAGGAGCGCGCGCGCGAGGCGCGAGGCGCGAGGCGCACTGAACT 373  
DB 1512 AGGTTCAAGGAGACAAACCCCGCTCCGAGAGGAGTGAAGAGACAGGGAACCTTTAATC 1571  
QY 374 TCCCCCAGATCACCTCTGTGCGCAGCGCCCTGCTGAGCATCAAGGTGGCGCGCCAGATCA 433  
DB 1572 TCCTTCAATCACTCTTTGGCAGCGACCCCTTGTCTCAATAAAGTGGGCGCCAGATAA 1631  
QY 434 AGGAGCCCTGTGACACCGCGCGCGAGCACACCGTGTCTGAGGAGATGAGCCTGCGCG 493  
DB 1632 AGGAAGCTCTCTTAGACACAGGAGCAGATGATACAGTATTAGAGAAAATAGCTTTGCCAG 1691  
QY 494 GCNAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAGGTGGCGCCAGTAGC 553  
DB 1692 GAAGTGGAAACCAAAATGATAGGAGGATTTGGAGGTTTATCAAGTAAAGCAGTAGT 1751  
QY 554 ACCAGATCCTGATCGAGATCTCGGCAAGAGGCCATCGGCAACCGTGTGATCGGCCCA 613  
DB 1752 ATCAATACTTATAGAAATTTGTGGAATAAGGCTATAGGTACAGTATTAGTAGGACCTA 1811  
QY 614 CCCCCGTGAACATCATCGSCCGCAACATGCTGACCCAGCTGGCTGCACTTCACTTCC 673  
DB 1812 CACCTGTCAACATAATTTGAAGAGATATGTTGACTCAGCTTGGATGCACTCTAAATTTTC 1871  
QY 674 CCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCGCATGGAAGCCCGCAAGG 733  
DB 1872 CAATTAGCCCATTGAACTGTACAGTAAATTTAAGCCAGGATGGATGGCCCAAGG 1931  
QY 734 TGAAGCAGTGGCCCTGACCGGAGAGAGATCAAGGCCCTGACCGCCATCTCGGAGGAGA 793  
DB 1932 TTAACAGTGGCCATTGACAGAGAGAAATTAAGCATTTACAGAAATTTGTAAAGAA 1991  
QY 794 TGAAGAGGAGGCAAGATCAACAGATCGGCCCGCGAGAACCCCTTACACACCCCGGTGT 853  
DB 1992 TGGAGAAGGAAGGAAAAATTAACAAAAATTTGGGCTTGAATAATCATATAACATCCAGTAT 2051  
QY 854 TCGCCCATCAAGAGAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCCGCAGCTCA 913  
DB 2052 TTGCCATAAAGAGAGAGCAGTACTTAAGTGGAGAAATTTAGTAGATTTTCAGGAGCTCA 2111  
QY 914 ACAAGCGCACCCAGGACTTCTGGAGTGCAGCTGGGCAATCCCGCAACCCCGCGGCGCTGA 973  
DB 2112 ATAAAGAACTCAAGACTTTTGGGAAGTTCAATTTAGGAATACCAACCCAGCAGGTTTAA 2171  
QY 974 AGAAGAGAGAGCGGTGACCGTCTGAGCTGGCGAGCGCCCTACTTCAGCTGCCCGTGG 1033  
DB 2172 AAAAGAAAAAATCAGTGACAGTACTGGATGTGGGGATGCATATTTTTCAGTTCCCTTAG 2231  
QY 1034 ACAGGAGCTTCGCGAAAGTACACCGCTTCCACATCCCGAGCATCAACAGAGACCCCGG 1093  
DB 2232 ATGAGGCTTCGCGAAATATACTGCAATTCACCATACCTAGTATAAACAATGAACACAG 2291

QY 1094 GCATCCGCTACAGTACAAAGTGTGCTGCTCCAGGGCTGGAAAGGCGACCCAGCACTTCC 1153  
 Db 2292 GGATTTAGATATCAATTAATATGTGCTTCCACAGGGATGGAAGGATCACCAGCAATATTC 2351  
 QY 1154 AGAGCAGCATGACCAAGATCCTGAGCCCTTCGCGCCGCGCAACCCGAGATCGTGATCT 1213  
 Db 2352 AGAGTAGCATGACAAATCTTAGAGCCCTTTAGGCGACGAATCCAAAATAGTCATCT 2411  
 QY 1214 ACCA-----GGCCCCCCTGTACGTGGGAGAGCACTTGGAGATCGGCCAGCACCGGCCA 1267  
 Db 2412 ATCAATATATATGATGACTGTATGTAGGGTCTGACTTTAGAAATAGGGCATCATAGAGCAA 2471  
 QY 1268 AGATCGAGGAGCTGCGCAAGCAGCTGCTGCGTGGGCTTCCACCCCGCGCAAGAAGC 1327  
 Db 2472 AATAGAGGAGTTAAGAGCACATCTATTAAAGTGGGATTCACCAACACAGATAGAAAC 2531  
 QY 1328 ACCAGAGAGCCCTTCTTCCGCAAT-----CGAGCTGCACCCCGCAAGTGGACCG 1381  
 Db 2532 ATCAGAAAGAACCCCAATTTCTTTGGATGGGGTATGAATCCATCTCGACAAATGGACAG 2591  
 QY 1382 TGAGCCCATCGAGCTGCGCGAGAGGAGAGCTGGACCGTGAAGCATCCAGAACCTGG 1441  
 Db 2592 TACAGCCCTAAGAGCTGCCAAGAAAGATAGTGGATGTCATATGATATACAGAACTTAG 2651  
 QY 1442 TGGCAAGCTGAATCGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1501  
 Db 2652 TGGAAATTTAAACTGGGCAAGTCAGATTTACCAGGGAATTAAGTGAGCAACTTTGTA 2711  
 QY 1502 AGCTGCTGGCGGCGCCCAAGCCCTGACCAACATCTGCGCCCTGACCGAGGAGCCGAGC 1561  
 Db 2712 AACTCCTTAGGGGGGCGCAAGAGCACTAACAGACATAGTACCACCTAACCTGAAGAGCAGAT 2771  
 QY 1562 TGGAGCTGGCGGAAACCGGAGATCTGCGGAGCCCGTGCACGGGTGTACTAGAAC 1621  
 Db 2772 TAGAATTAGCAGAAACAGGGAAATCTTAAAGAGCCAGTACATGGATATATTATGACC 2831  
 QY 1622 CCAGCAGGACCTGGTGGCGAGATCCAGAAAGCGGCCAGACCTAGTGGACCTACAGA 1681  
 Db 2832 CATCAAAAGACTTAATAGCTGAATACAGAAACAGGGGCGATGCCAATGGACATATCAA 2891  
 QY 1682 TCTACAGGAGCCCTTCAAGAACTCTGAAGACCGGCAAGTACGCAAGATCGCACCGCC 1741  
 Db 2892 TTTACCAAGAACCATTTCAAAATCTGAAACAGGGAAGTATGCAAAATGAGGACTGCTC 2951  
 QY 1742 ACACCAAGAGCTGAAGAGCTGACCGAGGCGGTGCGAAGATCGCCATGAGAGATCG 1801  
 Db 2952 ACACATATATGTATTAACAGTTAAACAGAGGCGAGTGCACCAAGATAGCCATAGAAAGCATAG 3011  
 QY 1802 TGATCTGGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGGAGACCTGGGAGACCT 1861  
 Db 3012 TAATAT--GGGAAAGACCCCTTAATTTAGACTTACCCATCCAAAAGAAACCTGGAGACAT 3070  
 QY 1862 GGTGAGCGACTACTGCGAGGCCACCTGGATPCCCGAGTGGAGTTTCGTAAACACCCCCC 1921  
 Db 3071 GGTGACAGACTATTGGCAGGCCACCTGGATTTCTGATTTGGAGTTTGTAAATACCCCTC 3130  
 QY 1922 CCTTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGCGCGAGACCTTCT 1981  
 Db 3131 CCTTAGTAAATTTATGTTACAGCTAGAAAGAACCCATAGTAGAGAGCAAACTTTCT 3190  
 QY 1982 ACGTGAGCGGCGCCCAACCGAGACCAAGATCGGCAAGCGCGGTACTGACCGAAC 2041  
 Db 3191 ATGTAGATGGAGCAGCTAATAGGGAATCTAAGTAGGAAAGCAGGGTATGTTACTGACA 3250  
 QY 2042 GGGCGCGCAGAGTCTGTGAGCTGACCGAGACCAACCAAGAGAGCGAGCTCGAG 2101  
 Db 3251 GAGGAAGCGCAAAATTTGTTCTTTAACTGAAACAAACAAATCAGAAAGCTGAATTTGAAG 3310  
 QY 2102 CCATCCAGCTGGCCCTGAGACAGCGGAGCGAGGTGAACATCGTACCGCAGCAGCGAGT 2161  
 Db 3311 CAATTCAGCTAGCTTTGCAAGATTCAGGAACAGAAAGTAACATAGTAACAGACTCAGT 3370  
 QY 2162 ACGCCCTGGGATCATTCAGGCCCGAGCCCGCAAGAGGAGAGCGAGCTGGTGAACAGA 2221

Db 3371 ATGCATTAGGAATCAATTAAGAGCACACCATGATTAAGTGAATCAGAGTTAGTCAACCAA 3430  
 QY 2222 TCATCGAGCAGCTGATCAAGAAGAGAGGTGTACCTGAGCTGGGTGCCCGCCCAAGG 2281  
 Db 3431 TAATAGAACAAATTAATAACAAAGAAAGAGTCTATCTGTCTATGGTACCAGCACATAAAG 3490  
 QY 2282 GCATCGGCGGCAACAGCAGATCGCAAGCTGTGAGCAGAGGCATCCGCAAGGTGCTGT 2341  
 Db 3491 GAATGGAGGGAATGAACAAAGTAGATAGATTAGTGAATTAGGAATAGGAAGTACTGT 3550  
 QY 2342 TCCTGGACGGCATCGAT 2358  
 Db 3551 TTCTAGATGGGATAGAT 3567

RESULT 14  
 US-09-872-733A-3  
 ; Sequence 3, Application US/09872733A  
 ; Patent No. 6656706  
 ; GENERAL INFORMATION:  
 ; APPLICANT: The Government of the United States of America, as  
 ; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND  
 ; TITLE OF INVENTION: SIV ENV GENES  
 ; FILE REFERENCE: 2026-4287051 HIV GAG/POL,SIV GAG & ENV  
 ; CURRENT APPLICATION NUMBER: US/09/872,733A  
 ; CURRENT FILING DATE: 2001-06-01  
 ; PRIOR APPLICATION NUMBER: PCT/US00/34985  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/173,036  
 ; PRIOR FILING DATE: 1999-12-23  
 ; NUMBER OF SEQ ID NOS: 19  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 3  
 ; LENGTH: 2467  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Mutated Human  
 ; OTHER INFORMATION: Immunodeficiency Virus - 1 Pol gene  
 US-09-872-733A-3

Query Match 46.3%; Score 1137; DB 4; Length 2467;  
 Best Local Similarity 83.4%; Pred. No. 8,6e-182;  
 Matches 1319; Conservative 0; Mismatches 250; Indels 12; Gaps 2;

QY 750 GAGATGGAGAGGAGGCGCAAGATCAACCAAGATCGGCCGCGAGAACCCCTCAACACCCCC 849  
 Db 7 GAGATGGAGAGGAGGAGGAGATCAGCAAGATCGGGCTGAGAACCCCTCAACACTCCA 66  
 QY 850 GTGTTCCGCTATCAAGAGAGGAGCAGCACCAAGTGGCGCAGCTGTGGACTTCCGCGAG 909  
 Db 67 GTCTTCGCAATCAAGAGAGGAGCAGTACCAAGTGGAGAAAGCTGGTGGACTTCAGAGAG 126  
 QY 910 CTGAACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGSCATCCCCCACCCTCCGCGG 969  
 Db 127 CTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAAGTGGGCTCCACATCCCGCTGGG 186  
 QY 970 CTGAAGAGAGAGAGAGCGTGCAGCTGTGACGTGGCGAGCGCTACTTTCAGCGTCCC 1029  
 Db 187 TTGAAGAGAGAGAGTGTGATGACAGTGTGATGATGGGTGATGCTTCTTCCGTTCCC 246  
 QY 1030 CTGACGAGGAGCTTCCGCAAGTACACCGCTTCCACCTCCCGAGCATCAACACAGAGAC 1089  
 Db 247 TTGGACGAGGAGCTTCAGGAAGTACATGCTGCTTCAGTACATCAACACAGAGACA 306  
 QY 1090 CCGGCGATCCGTACAGTACAAAGTGTGCGCCAGGCTGGAGGCGAGCCCGAGATC 1149  
 Db 307 CCAGGCGATCCGTACAGTACAAAGTGTGCGCCAGGAGTGAAGGGATACACGAGCATC 366  
 QY 1150 TTCCAGAGCAGCATCAACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGT 1209  
 Db 367 TTTCAGAGCAGCATCAACCAAGATCCTGGAGCCCTTCCGAGCAAAACCCAGACATCGT 426

1210 ATCTACCA-----GGCCCCCTGTACGTGGGAGGAGCCTGGAGATCGGCCAGACCGC 1263  
 Db ATCTATCAGTACATGAGACGACCTCTAGGTAGGAGTGAAGCTGGAGATCGGCCAGCAGG 486  
 1264 GCGAAGATCGAGAGCTGGGCAAGCAGCTGTCTGGCTGGGGCTTCAACACCCCGGAGCAG 1323  
 Db ACCAAGATCGAGAGCTGAGACAGCATCTGTTGAGGTGGGAGCTGACCAACCCAGCAG 546  
 1324 AAGCACCAGAGAGCGCCCTCTCTGCCCAT-----CGAGCTGCACCCCGACAGTGG 1377  
 Db AAGCACCAGAGAGCATCT 606  
 1378 ACCGTGAGCCATCGAGCTGCCGAGAGAGAGAGTGGACCGTGAACGACATCCAGAG 1437  
 Db ACAGTGCAGCCCATCGCTGCTGCTGAGAGGACAGCTGGACTGTGAACGACATACAGAG 666  
 1438 CTGGTGGCAAGCTGAAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1497  
 Db CTGGTGGGCAAGTGAAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 726  
 1498 TGCAAGCTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1557  
 Db TGCAAGCTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 786  
 1558 GAGCTGAGCTGGCGGAGAACCGGAGATCTCTGGGAGCGCCGTGACCGCGTGTACTAC 1617  
 Db GAGCTAGAACTGGCAGAGAACCGGAGATCTCTGAAGGAGCCAGTACATGAGTGTACTAC 846  
 1618 GACCCAGCAGGAGCTGTGGCGGAGATCCAGAGCAGGCGGCGGCGGCGGCGGCGGCGG 1677  
 Db GACCCAGCAGGAGCTGTGGCGGAGATCCAGAGCAGGCGGCGGCGGCGGCGGCGGCGG 906  
 1678 CAGATCTACAGGAGCGCTTCAAGAACCTGAAGACCGGCGGCGGCGGCGGCGGCGGCGG 1737  
 Db CAAATCTACAGGAGCGCTTCAAGAACCTGAAGACCGGCGGCGGCGGCGGCGGCGGCGG 966  
 1738 GCCCAGCAGGAGCTGTGGCGGAGATCCAGAGCAGGCGGCGGCGGCGGCGGCGGCGG 1797  
 Db GCCCAGCAGGAGCTGTGGCGGAGATCCAGAGCAGGCGGCGGCGGCGGCGGCGGCGG 1026  
 1798 ATCGTGTCTGGGCGGAGCGCCCAAGTTCGGCTGCGGCGGCGGCGGCGGCGGCGGCGG 1857  
 Db ATCGTGTCTGGGCGGAGCTCCCAAGTTCGGCTGCGGCGGCGGCGGCGGCGGCGGCGG 1086  
 1858 ACCTGGTGGAGCGGAGCTGTGGCGGAGATCCAGAGCAGGCGGCGGCGGCGGCGGCGG 1917  
 Db ACATGGTGGAGCGGAGCTGTGGCGGAGATCCAGAGCAGGCGGCGGCGGCGGCGGCGG 1146  
 1918 CCCCCCTGTGTGAGCTGTGTGATCCAGCTGGAGAGGAGCGCCATCATCGGCGGCGGAGC 1977  
 Db CCTCCCTGTGTGAGCTGTGTGATCCAGCTGGAGAGGAGCGCCATCATCGGCGGAGAGC 1206  
 1978 TTCTACGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2037  
 Db TTCTACGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1266  
 2038 GACCGGCGGCGGAGAGATCGTGGAGCTGACCGGAGCGGCGGCGGCGGCGGCGGCGGCGG 2097  
 Db AACCGAGGAGCAGAGAGTGGTGGAGCTGACCGGAGCGGCGGCGGCGGCGGCGGCGG 1326  
 2098 CAGGCGCATTCAGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2157  
 Db CAGGCGCATTCAGCTGGAGAGCGGAGCTGGAGAGTGAACATCGTGGAGAGAGCTCA 1386  
 2158 CAGTACGCGCTGGGAGATCATTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2217  
 Db CAGTACGCGCTGGGAGATCATTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1446  
 2218 CAGATCATCGAGCAGCTGATCAAGAGGAGAGGAGTGTACCTGAGCTGGGTCGCCCGCCAC 2277  
 Db CAGATCATCGAGCAGCTGATCAAGAGGAGAGGAGTGTACCTGAGCTGGGTCGCCCGCCAC 1506

2278 AAGGGCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2337  
 Db AAAGGAATTCGAGGAATGAACAAGTAGATAAAATATGTCAGTCTGGGATCCGGAAGGTG 1566  
 2338 CTGTTCTCTGACCGGATCGAT 2358  
 Db CTGTTCTCTGACCGGATCGAT 1587  
 RESULT 15  
 US-09-184-418C-4  
 ; Sequence 4, Application US/09184418C  
 ; Patent No. 6492110  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hahn, Beatrice  
 ; APPLICANT: Gao, Feng  
 ; APPLICANT: Shaw, George  
 ; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN  
 ; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1  
 ; FILE REFERENCE: D6287  
 ; CURRENT APPLICATION NUMBER: US/09/184,418C  
 ; CURRENT FILING DATE: 1999-11-02  
 ; NUMBER OF SEQ ID NOS: 112  
 ; SEQ ID NO 4  
 ; LENGTH: 8992  
 ; TYPE: DNA  
 ; ORGANISM: Human immunodeficiency virus type 1  
 ; FEATURES:  
 ; OTHER INFORMATION: isolates=92RW009; 139.1624:gag; 1690.4428:pol(N-terminus uncertain  
 ; OTHER INFORMATION: 4373.4951:vif; 4891.5181:vpr; 5162.7801:tat; 5301.7958:rev;  
 ; OTHER INFORMATION: 5403.5648:vpu; 5566.8148:env; 8150.8773:nef  
 ; US-09-184-418C-4  
 Query Match 45.5%; Score 1116.8; DB 4; Length 8992;  
 Best Local Similarity 68.3%; Pred. No. 2.4e-178;  
 Matches 1612; Conservative 0; Mismatches 732; Indels 16; Gaps 4;  
 14 TGCCCGAGCCCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 73  
 Db TGCCCGAGCCCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1280  
 74 AGGGCCCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 133  
 Db AGGGCCCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1340  
 1281 AGGGCCCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1400  
 134 ACTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 193  
 Db ATTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1400  
 194 AGGACTGCACTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 253  
 Db AAGACTGCACTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1459  
 254 AGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 311  
 Db AGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1519  
 312 -GCAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 370  
 Db GGNATGGGCGGAGAGATAGCTCTCTCTGTAAGAGGAGAGAGAGAGAGAGAGAGAGAG 1579  
 371 ACTTCCCCCAGATTCACCTCTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 430  
 Db ATTTCCTCTCAATCACTCTTTGGCAACGACCCCTTTGTCAAGTAAATATAGGAGGTGACG 1639  
 431 TCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 490  
 Db TAAGAGAGCTCTATTAGATACAGGAGAGATGATACAGTATTAGAGAGAAATAAATTTCG 1699  
 491 CCGGCAAGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 550  
 Db CAGGAAATGGAGAAACCAAAATGATAGGCGGAAATGGAGGTTTATCAAGGTAAGAGAGT 1759

QY 551 ACACACAGATCCTGATCGAGATCTGGGCAAGAGGCCATCGGCACCGTGTGATCGGCC 610  
 Db 1760 ATGATCAAAATCTATAGAAAATTTGTGAAAAAAGGCTATAGGTACAGTATTAGTAGGAC 1819  
 QY 611 CCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGCTGCACCCCTGAACT 670  
 Db 1820 CTACATCTGTCAACATAAATTTGGAGAAATATGTTGACCCAGATGGTGTACTTTAACT 1879  
 QY 671 TCCCATACGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCGGACGAGGCCCA 730  
 Db 1880 TTCCAAATAGTCTATTTAGAGCTGTACCACTAGCATTTAAAGCCAGGAATGGATGCCCAA 1939  
 QY 731 AGGTGAAGCAGTGGCCCTGACCGAGAGAGAAATCAAGCCCTGACCGCCATCTGCGAGG 790  
 Db 1940 AGGTTAAACAAATGCGCATTTGACAGAGAAATTAAGCAATTAAGAGAAATTTGTACAG 1999  
 QY 791 AGATGGAGAGAGGCGCAAGATCACCAAGATCGGCCCGGAGAACCCCTACAACCCCGCG 850  
 Db 2000 AAATGGAAAAAGAGGGAATAATTTCAAAAATCGGCGCTGAAAATCCATATAACACTCCAG 2059  
 QY 851 TGTTCGCCATCAGAGAGAGGAGCAGCACCAGTGGCGCAAGCTGGTGGACTTCCGCGAGC 910  
 Db 2060 TATTTGGCAATAAAAAGAGAGAGAGTACTAAGTGGAGAAATTAGTAGATTTTCAGGAAAC 2119  
 QY 911 TGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCATCCCCCACCOCGCGGCC 970  
 Db 2120 TCAACAAAGAACTCAAGACTTTTGGGAAGTCCAAATTAGGATACCAACCCAGCAGGCT 2179  
 QY 971 TGAAGAGAGAGAGCGTGCAGCTGTGACGTGGCGGACGCTACTTCAGCGTCCGCC 1030  
 Db 2180 TAAAGAGAGAAAAATCAGTGCAGTACTGGATGTGGGGGATGCATCTTCTCAGTTCCCT 2239  
 QY 1031 TGGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGCAGCATCAACAGAGACCC 1090  
 Db 2240 TAGATGAGAGCTTCAGGAATATACTGCATTCACCATACCTAGTATAAACAATGAACAC 2299  
 QY 1091 CCGGCATCCGCTACCACTACAAGCTGTGCTGCCCGAGGCTGGAAGGCGAGCCCGCAGCTCT 1150  
 Db 2300 CAGGAATTAGGTATCAATATAATGCTTCCACAGGGATGGAAGGATCACCCAGCAATAT 2359  
 QY 1151 TCCAGAGCAGCATCACCAAGATCTGTGAGCCCTTCGCGCCCGCAACCCCGAGATCGTGA 1210  
 Db 2360 TCCAAATAGTATGACAAATAATCTTAGAGCCCTTTAGGCAACAACCCAGAAATAGTGA 2419  
 QY 1211 TCTACCA-----GGCCCCCTGTACTGTGGGAGCGACCTGGAGATCGGCGAGACCGCG 1264  
 Db 2420 TCTATCAATATGATGATGCTGTACCTAGTACTGATCTAGAAATAGGGCAACATAGAG 2479  
 QY 1265 CCAAGATCGGAGCTGCGCAAGCAGCTGTGCGCTGGGGCTTCAACACCCCGCAAGA 1324  
 Db 2480 CAAAATAGAGAGTTAAGAGAACTTATTAAAGTGGGATTTACCAACCCAGACBAGA 2539  
 QY 1325 AGCACCAAGAGGAGCCCGCTTCTGCGCCAT-----CGAGCTGCACCCCGCAAGTGA 1378  
 Db 2540 AACATCAGAAAGAACCTCCATTTCTTGGATGGGGTATGAACCTTCATCTGACAAAATGGA 2599  
 QY 1379 CCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGACCGTGAACGACATCCAGAAAGC 1438  
 Db 2600 CAGTACAACTTATACGTGCCAGAAAGGATAGCTGGAATGTCATGATATACAGAGT 2659  
 QY 1439 TGGTGGGCAAGCTGAACCTGGGCGAGCCAGATCTACCCCGCATCAAGGTGGCCAGCTGT 1498  
 Db 2660 TAGTGGGHAATTTAACTGGGCAAGTCAAGTTTACCCAGGGGTTAAAGTAAGGCAATGT 2719  
 QY 1499 CCAAGCTCTGCGCGGCGGCGGAGCCCTGACCGACATGTCGCCCTGACCGAGGAGCGCG 1558  
 Db 2720 GTAAACTCTTTAGGGGAACCAAGAGCTTAAAGAGATAGTACCACTAACTGAAGAGAG 2779  
 QY 1559 AGCTGGAGCTGCCGAGAACCGGAGATCTTCCGCGAGCCCGTGCAGCGGCTGTACTACG 1618  
 Db 2780 AATTAGAAATTGGCAGAAACAGGGAATTTTAAAGAGAACCAAGTACATGGAGTATTATG 2839  
 QY 1619 ACCCAGCAAGGACCTGTGGGCGGAGATCCAGAGCAGGGCCACGACCAGTGGACCTACC 1678

Db 2840 ACCATCAAAAGACTTAATAGCTGAATACAGAAACAGGGCATGACCAATGACATATC 2899  
 QY 1679 AGATCTACAGGAGCCCTTCAAGACCTGAAGACCCGCGAAGTACGCCAAGATGCGCACCG 1738  
 Db 2900 AAATTTACCAAGAACCATTTCAAAAATCTGAAAACAGGAAAGTATGCAAAAAGAGAGACTG 2959  
 QY 1739 CCCACACCAAGCAGTGAAGCAGCTGACGAGCCCGTGCAGAAAGATCGCCATCGGAGAGA 1798  
 Db 2960 CCACACTAATGAGTTAAACAGTTAACAGAGCAGTGCAAAAGATAGCCATGGAAAGCA 3019  
 QY 1799 TCGTGTCTGGGCGAAGACCCCAAGTTTCGCGCTGCCATTCGAGAGAGACCTGGGAGA 1858  
 Db 3020 TAGTAATATGGGAAAGACTCTCTAAATTTAGATTATCCCATCCAGAAAGAAACATGGAAA 3079  
 QY 1859 CTTGGTGCAGCCGACTCTGCGAGCCACCTGGATCCCGAGTGGGAGTTTCGTGAACACCC 1918  
 Db 3080 CATGGTGCAGACACTATTGSCAAGCCACCTGGATTCCTGAGTGGAGTTTGTTAATACCC 3139  
 QY 1919 CCCCCCTGGTGAAGCTGTGTACAGCTGGAGAGAGCCCATCATCGGCGCGGAGACCT 1978  
 Db 3140 CTCCTCTAGTAAATATTATGTTACAGCTAGAGAAAGAACCCATATTAGGAGCAGAGACTT 3199  
 QY 1979 TCTAGCTGGAGCGCGCCCAACCGGAGACCAAGATCGGCAAGCGCGCTAGTGACCG 2038  
 Db 3200 TCTATGATGAGTGGAGCAGCTAATCGGAACTAAATAGAAAGCAGGCTATGTTACTG 3259  
 QY 2039 ACCGGGCGCGCAGAGATCGTGAAGCTGACCCGAGACCAACCCAGAGACCGAGCTGC 2098  
 Db 3260 ACAGAGAAAGCAGAAAATTTCTTCTTAACGAAAACAAATTCAGAAAGCTGAATTAC 3319  
 QY 2099 AGGCATCCAGCTGGCCCTGAGGACAGCGCAGAGGTGAACATCGTGCACCCAGAGCC 2158  
 Db 3320 AAGCAATTCAGCTAGCTTTACAGGATTCAGGATCAGAGATCAGAGCTCAATC 3379  
 QY 2159 AGTACGCCCTGGGCATCATCCAGCCCGCAGCCGCAAGAGAGAGAGCTGGTGAACC 2218  
 Db 3380 AGTATGCAATTAGGAATCATTTCAAGCAACACAGATAGCAGGATCGGAGCAGTCAATC 3439  
 QY 2219 AGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTACTGTGAGCTGGGTGCCGCCACA 2278  
 Db 3440 AAATAATAGAACAGTTAATAAAAAGGAAAGAGTCTACCTGTCTATGGGTACCAACATA 3499  
 QY 2279 AGGGCATCGGCGCAACGAGCAGATCGACAGCTGGTGAAGCAAGGGGATCCGCAAGTGC 2338  
 Db 3500 AAGGAATTTGGAGAAATGCAACAGTAGATAAATTAGTAAAGTAGTGAATCAGGAGAGTGC 3559  
 QY 2339 TGTTCCTCGAGCGCATCGAT 2358  
 Db 3560 TGTTCCTAGATGGAATAGAT 3579

Search completed: April 10, 2004, 16:22:57  
 Job time : 129.045 secs

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 06:34:42 ; Search time 4138 Seconds  
(without alignments)  
17774.420 Million cell updates/sec

Title: US-09-610-313-31  
Perfect score: 2463  
Sequence: 1 gtcagccaccatggccga.....gggctagccacgggtaattc 2463

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931050276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estmu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_htc:\*
  - 9: gb\_est1:\*
  - 10: gb\_est2:\*
  - 11: gb\_htc:\*
  - 12: gb\_est3:\*
  - 13: gb\_est4:\*
  - 14: gb\_est5:\*
  - 15: em\_estfun:\*
  - 16: em\_estom:\*
  - 17: em\_gss\_hum:\*
  - 18: em\_gss\_inv:\*
  - 19: em\_gss\_pln:\*
  - 20: em\_gss\_vit:\*
  - 21: em\_gss\_fun:\*
  - 22: em\_gss\_mam:\*
  - 23: em\_gss\_mus:\*
  - 24: em\_gss\_pro:\*
  - 25: em\_gss\_rod:\*
  - 26: em\_gss\_phg:\*
  - 27: em\_gss\_vrl:\*
  - 28: gb\_gss1:\*
  - 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	86.8	3.5	2598	11 AY103647	AY103647 Zea mays
2	85.6	3.5	951	12 BM321451	BM321451 rockefell
3	80.6	3.3	869	14 CK159167	CK159167 FGAS04056
4	80.2	3.3	1132	12 BM320864	BM320864 rockefell

5	79.6	3.2	1165	12 BM320900	BM320900 rockefell
6	77.9	3.2	867	12 BM321430	BM321430 rockefell
7	76.6	3.1	1550	12 BM321022	BM321022 rockefell
8	75	3.0	862	12 BM321023	BM321023 rockefell
9	74.4	3.0	914	28 BZ568300	BZ568300 pac82-164
10	73.2	3.0	853	12 BM321393	BM321393 rockefell
11	72.8	3.0	566	12 BM587428	BM587428 17006873
12	72.6	2.9	788	14 CB643171	CB643171 OGNJB003L
13	72.4	2.9	753	29 CC675888	CC675888 OGMCO51TH
14	72.2	2.9	640	10 BE601575	BE601575 HVSMEH009
15	72.2	2.9	688	14 CB648640	CB648640 OGNJB12C
16	72.2	2.9	764	14 CB651670	CB651670 OGNJB16L
17	72.2	2.9	766	14 CB642928	CB642928 OGNJB03F
18	72.2	2.9	809	14 CB641397	CB641397 OGNJB01A
19	72.2	2.9	841	14 CB651502	CB651502 OGNJB16H
20	72.2	2.9	851	29 CG250623	CG250623 OGMGE95TV
21	71.8	2.9	1933	11 AY325173	AY325173 Rattus no
22	71.6	2.9	500	12 BM372120	BM372120 EBR003_SQ
23	71.6	2.9	538	12 BM368580	BM368580 EBR008_SQ
24	71.6	2.9	540	9 AJ471121	AJ471121 AJ471121
25	71.6	2.9	566	13 BQ464692	BQ464692 HF02P20r
26	71.6	2.9	575	13 BU976068	BU976068 HA03B08r
27	71.6	2.9	576	13 BU984666	BU984666 HF04K03r
28	71.6	2.9	578	12 BM377112	BM377112 EBR005_SQ
29	71.6	2.9	579	10 BF253640	BF253640 HVSMEF000
30	71.6	2.9	582	12 BM372871	BM372871 EBR004_SQ
31	71.6	2.9	586	13 BU994273	BU994273 HM06H03r
32	71.6	2.9	588	13 BQ765498	BQ765498 EBR003_SQ
33	71.6	2.9	596	13 BU997745	BU997745 HI08O07r
34	71.6	2.9	600	9 AL508778	AL508778 AL508778
35	71.6	2.9	615	13 CA003718	CA003718 HU14O18r
36	71.6	2.9	621	13 BU986998	BU986998 HF13I08r
37	71.6	2.9	625	13 BQ768295	BQ768295 EBR008_SQ
38	71.6	2.9	637	13 BU999333	BU999333 HI14D01r
39	71.6	2.9	641	13 BU978712	BU978712 HA14B02r
40	71.6	2.9	645	13 BU977527	BU977527 HA11J06r
41	71.6	2.9	646	13 BU968956	BU968956 HB10B12r
42	71.6	2.9	655	13 BU996444	BU996444 HM13K04r
43	71.6	2.9	656	12 BM370222	BM370222 EBR008_SQ
44	71.6	2.9	659	13 CA031811	CA031811 HX11D06r
45	71.6	2.9	660	13 BQ762385	BQ762385 EBR001_SQ

ALIGNMENTS

RESULT 1	AY103647	2598 bp	mRNA	linear	HTC 16-OCT-2002
LOCUS	AY103647	AY103647	PCO142084	mRNA sequence.	
DEFINITION	Zea mays				
ACCESSION	AY103647				
VERSION	AY103647.1	GI:21206725			
KEYWORDS	HTC				
SOURCE	Zea mays				
ORGANISM	Zea mays				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.				
AUTHORS	Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whittitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.				
TITLE	Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes				
JOURNAL	Unpublished (2002)				
REFERENCE	2 (bases 1 to 2598)				
AUTHORS	Coe, E.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA				
COMMENT	If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the				



maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

FEATURES  
source

Location/Qualifiers  
1..2598  
/organism="Zea mays"  
/mol\_type="mrna"  
/db\_xref="MaizeDB:638378"  
/db\_xref="taxon:4577"  
/clone\_lib="Maize Mapping Project/DuPont Cornsensus Library"  
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

ORIGIN

Query Match 3.5%; Score 86.8; DB 11; Length 2598;  
Best Local Similarity 42.1%; Pred No. 0.59;  
Matches 813; Conservative 0; Mismatches 1102; Indels 15; Gaps 5;  
Qy 296 CCACACCGCGAGCTCAGAGTGGCGGACAAACCCCGCAGCGAGCGCGCGAGC 355  
Db |||||  
51 CCACACCTCCACTCTGCCACCGCGCGGCCCAACCAACACACCGCAGCGAGCAA 110  
Qy 356 GCCAGGACACCTGAATCTCCCGAGATCACTGTGGAGCGCCCTGTGAGCATCA 415  
Db |||||  
111 TGGCGGCAACGCGCCCATCTGGAGAGCAGCCGCTGAATCTGGGGCGCGCGCGCG 170  
Qy 416 AGGTGGCGGCGCAGATCAAGAGAGCCCTGTGTGACACCGCGCGCGAGCACCGTGTG 475  
Db |||||  
171 AGCTGGCGGAGGACCACTGGACGAGGTGAAGCGCATGTGTGGCGAGCGCCGCG 230  
Qy 476 AGGAGATGAGCTCCCGGCGAGTGGAGCCAGATGATCGCGGCGATCGCGGCTTCA 535  
Db |||||  
231 TGGTCAAGATCGAGGGTTCCACCCCTCCGCTGCGCCAGGTGGCGCGCTCGCCCA 290  
Qy 536 TCAAGGTGCGGCAGTACGACACAGATCTCGAGATCTCGGCAAGAGCCATCGGCA 595  
Db |||||  
291 AGGACGCTCGCGGTGCGCTGAGCTCGACGAGGAGGCCCGCCCGCGGTCAAGGCCA 350  
Qy 596 CCGTGTGATCGGCGCCACCCCGTGAACATCATCGCGCGGAGATGCTGACCGAGTGG 655  
Db |||||  
351 GCAGCGAGTGGATCTCGACTGTCATCGCCACCGCGCGCGACATCTACGGGTCAACACCG 410  
Qy 656 GCTGACCCCTGAATCTCCCATCAGCCCATCGAGACCGTGCCTGGAAGTGAAGCCCG 715  
Db |||||  
411 GTTGGGCGGACCTCCACCGCGCAGCAGGAGCGCGCGCTCCAGTCTGAGTGC 470  
Qy 716 GCATGAGCGGCCCCAAGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGA 775  
Db |||||  
471 TCAGGCATCTCAAGCGCGGATCTTCGCGCACCGGCGAGCGAGCGGACACGCTGCGCTCG 530  
Qy 776 CCGCATCTGAGGAGATGAGAGGAGGAGGAGGAGATCAACAGATCGGCGCCGAGAAC 835  
Db |||||  
531 AGGTCAACCGCGCGCGATGTGTGGCATCAACACCTCTCCAGGGTACTTCGGCA 590  
Qy 836 CCTCAACACACCCCGTGTTCGCGATCAAGAGAGAGGACAGCACCAAGTGGCGAAGTGG 895  
Db |||||  
591 TCCGCTTCAGATCTCTGAGGCCATCAAGAGTGTCTCAACACCGGTCTCAGGCCCTGCC 650  
Qy 896 TGGACTTCGCGAGTGAACAGCGCACCGAGACTTCTGGAGTGCAGTGGCATCC 955  
Db |||||  
651 TCGCGCTCCGGGCGACCATCAACCGGTGCGCGACCTGTGTCCTCTCTATCCCG 710  
Qy 956 CCCACCGCGCGGCTGAAGAGAGAGAGCGTGAACGCTGTGGAGTGGCGAGCGCT 1015  
Db |||||  
711 GCCTCATCAGGGCGCGCCCAACCGCAGGCGGTCAACG---TCGACGGGAGAGTGG 767  
Qy 1016 ACTTCAGGCTGCCCTGGACGAGGACTTCGCAAGTACACCGCTTCACTCCAGCA 1075  
Db |||||

Db 768 ACGCCGCGGAGGCGTTCAAGATCGCGGATCGAGGGCGCTTCTTCAAGCTCAACCCCA 827  
Qy 1076 TCAACAACAGAGACCCCGGATCGCTACAGTACAAAGTGTGCGCCAGAGGTGAAGG 1135  
Db AGGAGGCGCTCGCCATCGTCAACGGCACGTCTCGGCTCGGCGCTCGGCGCACCGTGA 887  
Qy 1136 GCAGCCCGACGATCTTCCAGAGAGATGACCAAGATCTCTGGAGCCCTTCCGCGCGCA 1195  
Db TGTACGACGCGCAACGCTCTGCGCTCTGTGCGAGGTCT---GTCCGCGTCTTCTGCG 944  
Qy 1196 ACCCGGAGATCGTATCTACAGGCCCCCTGTAGTGTGGCAGCGACTCTGGAGATCGGCC 1255  
Db AGGTCAATGAACCGCAAGCCCGAGTACACGGACACCTGACCCCAAGCTGAAGCACCA 1004  
Qy 1256 AGCACCGGCGCAAGATCGAGAGTGGCGAAGCACCTGCTGCGTGGGCTTCAACACCC 1315  
Db CGGGTTCATCGAGCGCGGCCCATCATGGAGACATCTCTGGATGGCAGTCTCTTATGA 1064  
Qy 1316 CCGACAGAGACACAGAGAGCCCCCTTCTGTGTGATGGGTACGAGTGCACCCCG 1375  
Db AGCAGGCCAAGAGTTGAACGAGCTGGACCGCTGTGAAGCCCAAGCAGGACAGGTACG 1124  
Qy 1376 ACAAGTGAACCGTGCAGCCCATCGAGCTGGCGCGAGGAGAGCTGACCGTGAACGACA 1435  
Db CGCTCCGACGCTCGCGCGAGTGGCTGGGCCCCAGATCGAGGT---ATCGCGCCGCCA 1181  
Qy 1436 TCCAGAGCTGTGGGCAAGCTGAATCTGGGCGACCCAGATCTACCCCGGATCAAGGTGC 1495  
Db CCAAGTCCATCGAGCGGAGGTCACTCCGTGAACGA---CAACCCGCTCATCGACGTC 1238  
Qy 1496 GCGAGCTGTGAAGTGTGCGCGCGCAAGGCCCTGACCGAGATCGTGCCTTCAACCG 1555  
Db ACCCGGCAAGCGCTGACCGCGCGCACTTCCAGGCGACCCCATCGGCGTGTCTATGG 1298  
Qy 1556 AGGAGCGGAGCTGAGCTGGCGAGAACCGGAGATCTTGGCGGAGCGCTGTGACGCGG 1615  
Db ACAAGCCCGCTCGCCATCGCCAAACATCGCAAGCTCATGTTGCGCAGTCTTCCGAGC 1358  
Qy 1616 TGTACTAGACCCCGCAAGGACCTGTGTGGCGAGATCCAGAGCAGGCGCAACCAAGT 1675  
Db TCGTCAACGAGTTTCAACAACCGGCTCACTCCAACTTGGCGGCGAGCGCGCAACCCCA 1418  
Qy 1676 GGAACCTACAGATCTTACAGGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAGA 1735  
Db GCCTGATCTAGGCTTCAAGGCGCAGAGATCGCCATGGCTCTTACTGTCCGAGTCC 1478  
Qy 1736 TCGCACCGCCCAACCAACGAGTGAAGAGTGAACGAGCGCGTGAAGAGATCGCCA 1795  
Db AGTACTTGGGCAACCCCATCAACCAACGCTGACAGCGCGGAGCAGCAACCAAGGAGC 1538  
Qy 1796 TGGAGAGCAT---CGTGATCTGGGCGAGACCCCAAGTTCCGCTTCCGCTTCCAGAGG 1852  
Db TGAATCTCCCTGGGCTCTGCTCGGCGAGAGACCGCGAGGCGATCGACATCTTGAAGC 1598  
Qy 1853 AGACCTGGGAGACCTGTGTGAGCCGACTACTGCGAGGCGCACTGATCCCCAGTGGAGT 1912  
Db TCATGTCTCCACTTACATCGTGGCGCTGTCCAGGCGCTGAGACCTTGGCGCCCTCGAGG 1658  
Qy 1913 TCGTGAACACCCCCCTTGTGTGAGTGTGTACAGCTGAGAGAGAGGCCCATCATCG 1972  
Db AGAATCAAGGCGTGGTGAAGAACCCGTGAACCGAGTGGCCAGAGAGTGTGTGACCA 1718  
Qy 1973 GCGCGGAGACCTTCTACGTGAGCGCGCGCAACCGCGAGACCAAGATCGGCAAGGCGC 2032  
Db TGAACCCCTCGGGGAGCTCTCCAGCGCCCGCTTTCAGCGAGAGAGTGTGATCAGCGCCA 1778  
Qy 2033 GGTAGTACCGACCGGGCGCGGAGAGATCTGTGAGCTGACCGAGACCAACCAAGCAGA 2092  
Db TCGACCGCGAGCGCTGTTCAGTACGCGGAGGAGCGCGCGCAGCGGAGCTTGGCGGTGA 1838  
Qy 2093 AGACCGAGCTGCGAGGCCATCCAGCTGGCCCTTGCAGGACAGCGGCGAGCGAGTGAACATCG 2152  
Db TGCAGAGCTGCGCGCGCTGCTGTGTGGACCAACCGCCCTCAGCAGCGGCGGACGCGAGCGG 1898

QY	2153	TGACCGACAGCAGTACGCTCGGCTATCCAGGCGCCAGCCGACAGAGCGAGCG	2212	
Db	1899	AGCCCTCCGTGTTCTCCAGATCACAGGTTGAGGAGGAGCTCCGCGGTCTCCCTC	1958	
QY	2213	AGCTGGTGAA	2222	
Db	1959	AGGAGGTGA	1968	
RESULT 2				
LOCUS	BM321451	951 bp	mRNA	linear
DEFINITION	rockefeller.0.1211 Mastigamoeba balamuthi lambda ZAP II Library			EST 03-JAN-2002
	Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC			
	3.3.1.1), mRNA sequence.			
ACCESSION	BM321451			
VERSION	BM321451.1	GI:18055857		
KEYWORDS	EST.			
SOURCE	Mastigamoeba balamuthi			
ORGANISM	Mastigamoeba balamuthi			
REFERENCE	1 (bases 1 to 951)			
AUTHORS	Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.			
	Baptiste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W.,			
	Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and			
	Philippe, H.			
TITLE	The analysis of 100 genes supports the grouping of three highly			
	divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)			
MEDLINE	21819461			
PUBMED	11930664			
COMMENT	Contact: Muller Miklos			
	Laboratory of Biochemical Parasitology			
	The Rockefeller University			
	1230 York Avenue, New York, NY 10021, USA			
	Email: mmuller@rockvax.rockefeller.edu			
	Insert Length: 951 Std Error: 0.00			
	POLYA-No.			
FEATURES				
source	1..951			
	/organism="Mastigamoeba balamuthi"			
	/mol_type="mRNA"			
	/strain="ATCC 30984"			
	/db_xref="taxon:108607"			
	/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"			
	/note="syn: Phreatamoeba balamuthi"			
ORIGIN				
	Query Match	3.5%;	Score 85.6;	DB 12; Length 951;
	Best Local Similarity	46.0%;	Pred No. 0.68;	
	Matches	323;	Conservative	0; Mismatches 376; Indels 3; Gaps 1;
QY	1728	CGCCAAAGATCGGACCGCCACACCAACGACGTGAAGCAGTGCACCGAGGCGCGTGCAGAA	1787	
Db	219	CGCCAAAGATCGGCTGGTCTCGTCAACATCTTCTCGACGAGGACCAACGCGCGCGCGC	278	
QY	1788	GATCCCATGGAGACATCGTATCTGGGCGAGACCCCAAGTTCGCGTCCCATCCA	1847	
Db	279	CATCCGCGAGCGCGCGTCTCGGTCTTCGCTTGAAGGCGGAGAACCTCCAGGAGTACTG	338	
QY	1848	GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTCTGGCAGGCGCACCTGGATCCCGCGAGTG	1907	
Db	339	GGAGTGACCTGAAGGCGCTGTCTTCGCGCCCTACCAAGGCGCTCAGATCATCTGTCGA	398	
QY	1908	GGAGTTGTTGAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGGAGAGAGGCCAT	1967	
Db	399	CGACGCGGTGACGCGACTCTGATGATCCACAAGGCGGTTCGCGCGCGAGGCAACCCCAA	458	
QY	1968	CATCGCGCGGAGACCTTCTAGCTGGAACGCGCGCGCCGACACCGGAGACCAAGATCGGAA	2027	
Db	459	GCTGCTGGAGGACGACGAGGCGCTCGAGGAGTTCGCTTCCTCACACAGCTGCTCAAGCA	518	
QY	2028	GCGCGGCTACGTGACCGACCGCGCGCGCGGAGAAAGATCGTGAACCTGACCGAGACCAACAA	2087	

Db	519	GGTCCAGAAGGAGCAGCGCGCTTCTGGCACAAGATCTCCCGAGATCCCGCGGTGTGAG	578	
QY	2088	CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTCAGACAGAGCGGCGAGCGTGAA	2147	
Db	579	CGAGGAGACGAGCTGGCTGACAGCTGACAGCTGCACCGCGCAGCGAGCTGCT	638	
QY	2148	CATCGTGACCGACAGCCAGTACGCGCTGGGGATCATCAGGCCCGAGCCCGACAGAGCGGA	2207	
Db	639	GTTCCCGCGCTCAACGTCACAGAC---TCTNTCCCAAGAGCAAGTTCGACAACATCTA	695	
QY	2208	GAGCGAGCTGGTGAACACAGATCATCGAGCAGCTGATCAAGAAGGAGAAGTGTACTCTGAG	2267	
Db	696	CGGCTGCGCCACTCGCTCATCGAGCGCATCAAGCGCGGACCGCTGATGCTCGGCGG	755	
QY	2268	TTGGGTGCGCGCCACAAAGGCGATCGCGGCAAGAGATCGACAGCTGTGTGAGCAA	2327	
Db	756	CAAGGTGCGCGCTCGCTCGCGGCTACGCGGACGCTGCGGCAAGGCGCTCGCGCGTCTGCG	815	
QY	2328	GGGATCGCGCAAGTGTCTTCTGGAGCGCATGATGGCGGCATCGTGTACTACAGTA	2387	
Db	816	CGCGCAGGCGTCCCGCTCATCGTACGAGATGACCCCATCTGCGCGCTGCAGGCGTC	875	
QY	2388	CATGACGACCTGTACGTGGCGCGCGCGCTAGGATCGA	2429	
Db	876	GATGCGCGCTTCGAGGTCAACACGCTCGAGCGCGGCTCGA	917	
RESULT 3				
LOCUS	CK159167/c	869 bp	mRNA	linear
DEFINITION	FGAS040564 Triticum aestivum FGAS: TaLts Triticum aestivum cDNA,			EST 05-DEC-2003
	mRNA sequence.			
ACCESSION	CK159167			
VERSION	CK159167.1	GI:38985053		
KEYWORDS	EST.			
SOURCE	Triticum aestivum (bread wheat)			
ORGANISM	Triticum aestivum			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
	Poideae; Triticeae; Triticum.			
	1 (bases 1 to 869)			
	Allard, F., Crosby, W. I., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,			
	Genswein, B., Graf, R., Gulick, P., Hrycan, L. D., Larocque, A.,			
	Links, M. G., McCarthy, E. L., Monroy, A., Muzak, I., Nilsson, D.,			
	Penniket, C., Roach, J. L. and Sarhan, F.			
	Functional Genomics of Abiotic Stress In Wheat and Canola Crops			
	Unpublished (2003)			
TITLE	Unpublished (2003)			
JOURNAL	Contact: Wm L Crosby			
COMMENT	Bioinformatics			
	University of Saskatchewan, Department of Computer Science			
	1C101 Engineering Building, 57 Campus Drive, Saskatoon,			
	Saskatchewan, S7N 5A9, Canada			
	Tel: 306 966 1769			
	Fax: 306 966 2033			
	Email: fgas.est@cs.usask.ca			
	This sequence is the direct result of the Base calling software			
	Phred (default parameters). It is the raw base calls. To aid in the			
	identification of the high quality insert the software Lucy			
	(default parameters) has been run on this sequence. Lucy identified			
	the region [126,636].			
	Plate: TaLts37 row: N column: 23.			
FEATURES				
source	1..869			
	/organism="Triticum aestivum"			
	/mol_type="mRNA"			
	/cultivar="wheat line PI 178383"			
	/db_xref="taxon:4565"			
	/lab_host="DHS alpha"			
	/clone_lib="Triticum aestivum FGAS: TaLts"			
	/notes="Organ: Crown; Vector: pGEM-T; SSH (suppression			
	subtractive hybridization) cDNA library from genotype			
	PI178383 cold hardened at 2 C for 21 days and 49 days			

(equal amount of cDNA pooled together before subtraction, tester) and subtracted against genotype Norstar cold hardened at 2 °C for 1 day (24 h) (driver). Modified Smart cDNA (Clontech) priming and non-directional cloning<sup>19</sup>

## ORIGIN

Query Match	3.3%	Score 80.6	DB 14	Length 869
Best Local Similarity	44.9%	Pred. No. 2.7		
Matches 305	Conservative 0	Mismatches 374	Indels 0	Gaps 0
QY	580	AAGAGGCGCATCGGACCGTGCTGATCGGCCCCACCCGCTGAACATCATCGGCGCGCAAC	639	
Db	827	ATGCGGCGCGCCGCAACAACGAGGATCAACACCGAGGCGAGNAACAACCAACCAACAC	768	
QY	640	ATGCTGACCGAGCTGGGCTGCACCTGAAATTCCCATCAGGCCCATCGAGACCGTGCC	699	
Db	767	CACAACGACGACAACAGCGACCAACACCAACCAACCAACCAACCAACCAACCAAC	708	
QY	700	GTGAGCTGAAGCCCGGCGATCGAGGGCCCCAAGGTGAGCGTGGGCCCTGACGGAGG	759	
Db	707	ACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC	648	
QY	760	AGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAGGGCGAAGATCAACCAAG	819	
Db	647	AACAACAACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC	588	
QY	820	ATCGGCCCGGAGAACCCCTACAACACCCCGCTGTTGCCATCAAGAAGAAGGACGACCC	879	
Db	587	AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC	528	
QY	880	AAGTGGCGCAGCTGGTGGACTTCCGGAGCTGAAACAAGCGCACCGAGCTTTCTGGGAG	939	
Db	527	AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC	468	
QY	940	GTGCACTGGCGCATCTCCCAACCCCGCGGCTGTAAGAAGAAGAAGGCGTACCGTGCTG	999	
Db	467	AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC	408	
QY	1000	GACGTGGCGAGCGCTACTTTCAGCTGCCCTCGACGAGGATTCGCAAGTACACCGCC	1059	
Db	407	GACAACAACGACAACAACGACAACAACAACAACAACAACAACAACAACAACAACAAC	348	
QY	1060	TTCAACATCCCGACGATCAACAAGAGACCCCGGATCGGTACAGTACAAAGCTGCTG	1119	
Db	347	GAACCAACGACACGACACACAACAACAACAACAACAACAACAACAACAACAACAAC	288	
QY	1120	CCCAAGGCTGGAGGGCAGCCCAAGCATTTCCAGACGAGCATCAACCAAGATCTCGAG	1179	
Db	287	AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC	228	
QY	1180	CCCTTCGCGCCCGCAACCCGAGATCTGTATCTACCGGCCCTGTCTGCTGGCGACG	1239	
Db	227	AACAACAACAACAACAACGACGACGACGCTACGATTAACAACAACAACAACAACAAC	168	
QY	1240	GAQCTGGAGATCGGCCAGC	1258	
Db	167	AACGGCAACAACAACAAC	149	

RESULT 4	BM320864	1132 bp	mRNA	linear	EST 03-JAN-2002
LOCUS	BM320864				
DEFINITION	Rockefeller.0.46 Mastigamoeba balamuthi lambda ZAP II Library				
	Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA				
	sequence.				
ACCESSION	BM320864				
VERSION	BM320864.1	GI:16055270			
KEYWORDS	EST.				
SOURCE	Mastigamoeba balamuthi				
ORGANISM	Mastigamoeba balamuthi				
	Eukaryota; Pelobiontida;				
REFERENCE	1 (bases 1 to 1132)				
AUTHORS	Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.				

Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and Philippe, H.  
The analysis of 100 genes supports the grouping of three highly  
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba  
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
21819451  
11830664

CONTACT: Muller Miklos  
Laboratory of Biochemical Parasitology  
The Rockefeller University  
1230 York Avenue, New York, NY 10021, USA  
Email: mmuller@rockvax.rockefeller.edu  
Insert Length: 1132 Std Error: 0.00

FEATURES	SOURCE
1. The first two columns are labeled "FEATURES" and "SOURCE".	
2. The third column is labeled "FEATURES".	
3. The fourth column is labeled "SOURCE".	
4. The fifth column is labeled "FEATURES".	
5. The sixth column is labeled "SOURCE".	
6. The seventh column is labeled "FEATURES".	
7. The eighth column is labeled "SOURCE".	
8. The ninth column is labeled "FEATURES".	
9. The tenth column is labeled "SOURCE".	
10. The eleventh column is labeled "FEATURES".	
11. The twelfth column is labeled "SOURCE".	
12. The thirteenth column is labeled "FEATURES".	
13. The fourteenth column is labeled "SOURCE".	
14. The fifteenth column is labeled "FEATURES".	
15. The sixteenth column is labeled "SOURCE".	
16. The seventeenth column is labeled "FEATURES".	
17. The eighteenth column is labeled "SOURCE".	
18. The nineteenth column is labeled "FEATURES".	
19. The twentieth column is labeled "SOURCE".	
20. The twenty-first column is labeled "FEATURES".	
21. The twenty-second column is labeled "SOURCE".	
22. The twenty-third column is labeled "FEATURES".	
23. The twenty-fourth column is labeled "SOURCE".	
24. The twenty-fifth column is labeled "FEATURES".	
25. The twenty-sixth column is labeled "SOURCE".	
26. The twenty-seventh column is labeled "FEATURES".	
27. The twenty-eighth column is labeled "SOURCE".	
28. The twenty-ninth column is labeled "FEATURES".	
29. The thirtieth column is labeled "SOURCE".	
30. The thirty-first column is labeled "FEATURES".	
31. The thirty-second column is labeled "SOURCE".	
32. The thirty-third column is labeled "FEATURES".	
33. The thirty-fourth column is labeled "SOURCE".	
34. The thirty-fifth column is labeled "FEATURES".	
35. The thirty-sixth column is labeled "SOURCE".	
36. The thirty-seventh column is labeled "FEATURES".	
37. The thirty-eighth column is labeled "SOURCE".	
38. The thirty-ninth column is labeled "FEATURES".	
39. The fortieth column is labeled "SOURCE".	
40. The forty-first column is labeled "FEATURES".	
41. The forty-second column is labeled "SOURCE".	
42. The forty-third column is labeled "FEATURES".	
43. The forty-fourth column is labeled "SOURCE".	
44. The forty-fifth column is labeled "FEATURES".	
45. The forty-sixth column is labeled "SOURCE".	
46. The forty-seventh column is labeled "FEATURES".	
47. The forty-eighth column is labeled "SOURCE".	
48. The forty-ninth column is labeled "FEATURES".	
49. The fiftieth column is labeled "SOURCE".	
50. The fifty-first column is labeled "FEATURES".	
51. The fifty-second column is labeled "SOURCE".	
52. The fifty-third column is labeled "FEATURES".	
53. The fifty-fourth column is labeled "SOURCE".	
54. The fifty-fifth column is labeled "FEATURES".	
55. The fifty-sixth column is labeled "SOURCE".	
56. The fifty-seventh column is labeled "FEATURES".	
57. The fifty-eighth column is labeled "SOURCE".	
58. The fifty-ninth column is labeled "FEATURES".	
59. The sixtieth column is labeled "SOURCE".	
60. The sixty-first column is labeled "FEATURES".	
61. The sixty-second column is labeled "SOURCE".	
62. The sixty-third column is labeled "FEATURES".	
63. The sixty-fourth column is labeled "SOURCE".	
64. The sixty-fifth column is labeled "FEATURES".	
65. The sixty-sixth column is labeled "SOURCE".	
66. The sixty-seventh column is labeled "FEATURES".	
67. The sixty-eighth column is labeled "SOURCE".	
68. The sixty-ninth column is labeled "FEATURES".	
69. The seventieth column is labeled "SOURCE".	
70. The seventy-first column is labeled "FEATURES".	
71. The seventy-second column is labeled "SOURCE".	
72. The seventy-third column is labeled "FEATURES".	
73. The seventy-fourth column is labeled "SOURCE".	
74. The seventy-fifth column is labeled "FEATURES".	
75. The seventy-sixth column is labeled "SOURCE".	
76. The seventy-seventh column is labeled "FEATURES".	
77. The seventy-eighth column is labeled "SOURCE".	
78. The seventy-ninth column is labeled "FEATURES".	
79. The eightieth column is labeled "SOURCE".	
80. The eighty-first column is labeled "FEATURES".	
81. The eighty-second column is labeled "SOURCE".	
82. The eighty-third column is labeled "FEATURES".	
83. The eighty-fourth column is labeled "SOURCE".	
84. The eighty-fifth column is labeled "FEATURES".	
85. The eighty-sixth column is labeled "SOURCE".	
86. The eighty-seventh column is labeled "FEATURES".	
87. The eighty-eighth column is labeled "SOURCE".	
88. The eighty-ninth column is labeled "FEATURES".	
89. The ninetieth column is labeled "SOURCE".	
90. The ninety-first column is labeled "FEATURES".	
91. The ninety-second column is labeled "SOURCE".	
92. The ninety-third column is labeled "FEATURES".	
93. The ninety-fourth column is labeled "SOURCE".	
94. The ninety-fifth column is labeled "FEATURES".	
95. The ninety-sixth column is labeled "SOURCE".	
96. The ninety-seventh column is labeled "FEATURES".	
97. The ninety-eighth column is labeled "SOURCE".	
98. The ninety-ninth column is labeled "FEATURES".	
99. The hundredth column is labeled "SOURCE".	

```

FEATURES
  source
    1, 1132
    location/Qualifiers
      /organism="Mastigamoeba balamuthi"
      /mol_type="mrna"
      /strain="ATCC 30984"
      /db_xref="taxon:108607"
      /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
      /note="Syn. Phreatamoeba balamuthi"

```

## ORIGIN

Query Match	3.3%; Score 80.2; DB 12; Length 1132;
Best Local Similarity	45.3%; Prid. No. 3.1.1;
Matches 375; Conservative	0; Mismatches 444; Indels 9; Gaps 2;
QY	139 CGCGCCCCCACAAGAGGCGTCTCGAAGTGCGCAGAGGGGCCACAGATGAAGAC 198
Db	79 CGCGCGAGGCAAGACGACTACCGCGCGCANCTGGTGATCCAGGACAGAACAAG 138
QY	199 TGCACCGAGCGCAGGCCAACTTTCTTCGCGAGGAGCTGGCTTCCCGCAGGCAAGGCC 258
Db	139 TACAACAGCCCAAGTACCGCTTCGTC-----GTCCGTTCAACAAGGACATGTC 192
QY	259 CGCGAGTTCCCGACGAGCAGAAACCGCGCAACAGCCCAACAGCGCGAGCTCGAGGTG 318
Db	193 TGCCAGATCGCTACGCCCAAGATCGACGCGGCCACCATCTCTCGCGCGCGCTACTCGAC 252
QY	319 CGCGCGCAACACCCCGCACGAGGCGCGCGGCGAGCGCGAGGCGCACCTGAACTTCCCC 378
Db	253 GAGCTCACCGCTTCGCGGTCAAGCTCGGCTGACCAACTACGCGCGCGCTACGCGACT 312
QY	379 CAGATCACTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGGGCGGCAGATCAAGGAG 438
Db	313 GGCCTGTCTGTGGCCCCCGTGTCTGAAGAAGCTCAACCTCGACTCCCAAGTACGAGGT 372
QY	439 GCCCTGTCTGACACCGCGCGCCGACGACCGCTCTGTGAGGAGATGAGCTCCCGCGCAAG 498
Db	373 GTCAAGAAGGTCAACCGCGAGGACTCAACCTCGAGGAGCTCGAGACGGGCCCCCGCGT 432
QY	499 TGGAAAGCCCAAGATGATCGCGCGCATCGCGCGTTCATCAAGGTGCGCCAGTACGACAG 558
Db	433 TTCAAGGCCCTGTCTCGAGCTCGGCTGTGTCCGCACTCGACTGGCGCGCGGTTCGCC 492
QY	559 ATCTGTATCGAGATCTCGGCAAGAAGGCCATCGGACCGTGCTGATCGGCCCCACCCCC 618
Db	493 GCGCTCAAGGCGATGTTCGACGCGCGCGTCAACGTCCCC---CACAGCGAGACCGGTTTC 549
QY	619 GTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTCCCCATC 678
Db	550 GTGCGCTTCAACGCGCAACAAGAAGGAGCTCAACGCGCGGTCCTCCGCAAGTACATCTTC 609
QY	679 AGCCCCATCGAGACCGTGCCGTGTAAGCTGAAGCCGCGCATGGACGCGCCCAAGGTGAAG 738
Db	610 GCGCGCCACGTGCGCGGTACATGAAGCTTCTCAAGGAGCAGGACGCGCGCCTTCGAC 669
QY	739 CAGTGGCCCCGTGACCGAGGAGAGATCAAGGCCCTTGAACCGCATCTGCGAGGAGATGAG 798
Db	670 GCGCAGTTCTCGCGCTACGCCAAGGAGGGTGTCAACGCGCATGCTCTGAGAAGATCTAC 729
QY	799 AAGGAGGCGCAAGATCAACCAAGATCGCGCCCGGAGAACCCCTACAAACCCCCGTGTGCC 858

Db 730 ACCGAGCCCAACAGCAGATCCGCGCCGACCGACCTTCGTCCCAAGCGCGCTCGAAG 789

Qy 859 ATCAAGAAGAGGACAGACCAAGTGGCGGCAAGTGGTGGAGCTTCGCGAGAGTGAACAAG 918

Db 790 CCCGAGGCGGCAAGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACCAAGCGGCAAG 849

Qy 919 CGACCCAGGACTTCTGGAGGTGACGCTGGGCACTCCCAACCCCGCC 966

Db 850 AACCGCTGCGCCAGAGAAGTGGCTGGGCTACCCCGAGCGCCCC 897

RESULT 5  
BM320900 1165 bp mRNA linear EST 03-JAN-2002  
LOCUS rockefeller.0.1153 Mastigamoeba balamuthi lambda ZAP II Library  
DEFINITION Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA sequence.

ACCESSION BM320900  
VERSION BM320900.1 GI:18055306  
KEYWORDS Mastigamoeba balamuthi  
SOURCE Mastigamoeba balamuthi  
ORGANISM Mastigamoebidae; Mastigamoeba.

REFERENCE 1 (bases 1 to 1165)  
AUTHORS Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and Philippe, H.  
The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba  
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)

TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)

MEDLINE 21819461

PUBMED 11830664

COMMENT Contact: Muller Miklos  
Laboratory of Biochemical Parasitology  
The Rockefeller University  
1230 York Avenue, New York, NY 10021, USA  
Email: mmuller@rockvax.rockefeller.edu  
Insert Length: 1165 Std Error: 0.00  
POLYA=No.

FEATURES  
Location/Qualifiers  
1..1165  
/organism="Mastigamoeba balamuthi"  
/mol\_type="mRNA"  
/strain="ATCC 30984"  
/db\_xref="taxon:108607"  
/clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
/note="syn: Phreatamoeba balamuthi"

ORIGIN  
Query Match 3.2%; Score 79.6; DB 12; Length 1165;  
Best Local Similarity 45.1%; Pred. No. 3.7; Mismatches 506; Indels 16; Gaps 3;  
Matches 428; Conservative 0

Qy 90 CATCAAGTGTCTCAACTGGCGCAAGGAGGCGCCACATCGCCCGCAACTGCGCGCCCCCG 149

Db 46 CGTCAAGACCAAGCGCTACTTCAAGCGCTTCCAGACCCAGTTCGCTGCGCGCGGAGGG 105

Qy 150 CAAGAAGGGGTGTGAAGTGGCGCAAGAGGCGCCACAGATCAAGGACTGCACCGAGCG 209

Db 106 CAAGACGGGACTACCGCGCGCGCCANCTGGTGTATCCAGGACAGAACTACAGCCCC 165

Qy 210 CCAGGCGCAACTTCTTCGCGAGGACTGGCTTCCCGCGAGGCAAGCGCCGCGAGTTCCC 269

Db 166 CAAGTACCGTTTGTC-----GTCCGCTTCAACACAGGACATCGTTCGCAGATCGC 219

Qy 270 CAGCGAGCAGAACCGGCGCAACAGCCCCACAGCGCGGAGCTCAGGTGGCGGGGACAA 329

Db 220 CTACGCCAAGATCGACCGGCGCCACATCTCTCGCGCGCGCTACTCGCAGAGTCAACCG 279

Qy 330 CCCCAGGAGGCGCGCGCGCGAGCGCCAGGCGACCTGAATTCCTCCCGAGATCACCT 389

Db 280 CTTGGCGGTCAAGCTCGGCTGACCAACTAGCGCGCGCGCTACGCGACTGGCTGTGCT 339

Qy 390 GTGGCAGCGCCCTGTGTGAGCATCAAGGTGGGCGGCGCAGATCAAGAGGCGCTGTCTGGA 449

Db 340 GSCCGCGCTGTGTGAAGAAGCTCAACCTCTACTTCCAAAGTACGAGGGTGTCAAGAAGGT 399

Qy 450 CACCGCGCGCCAGCAGACCGCTGTGAGGAGATGAGCTGCCCGCAAGTGGAGCCCAA 509

Db 400 CAAACGGCGAGGACTACAAGCTGAGAGGCTCGACACGGGCGCCCGCTTCAAGGSCC- 457

Qy 510 GATGATCGGCGGCATCGGCGGCTTCAACAGGTGGCGGCGAGTACGACAGAGTCTCTATCGA 569

Db 458 --TGCTCGAGCTCGGCTGTTCGCGACCTCGACTGCGCGCGCGCTGTTTCGCGCGCTCAA 515

Qy 570 GATCTCGGCAAGAAGGCCATCGGCAACGCTGATCGGCGCCCGCCACCGCTGAACATCAT 629

Db 516 GGGCATGTTCGAGCGCGGCTCAACGTCCCCACAGCAGAGCCCGCTTCGTCGGCTTCAA 575

Qy 630 CGCGCGCAACATGCTGACCCAGCTGGGTGACCTGAACTTCCCATCAGCCCCATCGA 689

Db 576 CGCGGACAGAAGGAGCTCAACGCGCGGTCTCTCGCAAGTACATCTTCGCGGCGCACGT 635

Qy 690 GACCGTGGCGGTGAAGTGAAGCCCGGCATGACCGGCCCAAGGTGAACAGTGGCGCCCT 749

Db 636 CGCGCGGTATCAAGAGCTTCTCAAGAGCAGAGCGCGCGCTTCGACCGCCAGTTCTC 695

Qy 750 GACCGAGGAGAGATCAAGGCGCTGACCGCCCTCTCGAGGAGATGGAGAGAGGCGCAA 809

Db 696 GCGCTAGCCNAGAGGGTGTACCGCGGACATGCTCGAGAAGATCTACACGAGGCCCCA 755

Qy 810 GATCACAGATCGGCGCGCGAGAACCCCTACACACCCCGCTTCGCCCATCAAGAAGAA 869

Db 756 -----CAAGCAGATCCGCGCGAGCCGACCTTGTTCCTCAAGCGCGCTTCGAGCCCGA 809

Qy 870 GGACAGCACCAAGTGGCGCAAGCTGTGTGACTTTCGCGAGCTGAACAAGCGCACCCAGA 929

Db 810 GGGCGCCAGCCCAAGCACTTGGGGCAAGCGCAGGCTGACGTACCAAGCGCAAGAAGCG 869

Qy 930 CTTCTGGAGGTGACGTGGGCAATCCCCACCGCGCGCTTCGAGAGAGAGAGCGCT 989

Db 870 CGTGGCGCGAGAGAGGTTCGCGCTGGGCTACCGCGCGCGCGCGCGCGCGCGCGCGCG 929

Qy 990 GACCGTGTGAGTGGGAGCGGCTTACCTGAGCGTGGCGCTTCGAGCGAG 1039

Db 930 CGGTGCGCTGTACACCGCCATCTCCCGCTGCGCGGTGTGCTGTCTGCGG 979

RESULT 6  
BM321430 867 bp mRNA linear EST 03-JAN-2002  
LOCUS rockefeller.0.1153 Mastigamoeba balamuthi lambda ZAP II Library  
DEFINITION Mastigamoeba balamuthi cDNA similar to ribosomal protein S4, mRNA sequence.

ACCESSION BM321430  
VERSION BM321430.1 GI:18055836  
KEYWORDS Mastigamoeba balamuthi  
SOURCE Mastigamoeba balamuthi  
ORGANISM Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.

REFERENCE 1 (bases 1 to 867)  
AUTHORS Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and Philippe, H.  
The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba  
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)

TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)

MEDLINE 21819461

PUBMED 11830664

COMMENT Contact: Muller Miklos  
Laboratory of Biochemical Parasitology  
The Rockefeller University  
1230 York Avenue, New York, NY 10021, USA  
Email: mmuller@rockvax.rockefeller.edu  
Insert Length: 867 Std Error: 0.00

**SOURCE** Mastigamoeba balamuthi  
**ORGANISM** Mastigamoeba balamuthi  
**REFERENCE** Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.  
**AUTHORS** 1 (bases 1 to 1550)  
 Baptiste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W.,  
 Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and  
 Philippe, H.  
**TITLE** The analysis of 100 genes supports the grouping of three highly  
 divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba  
 Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
**JOURNAL**  
**MEDLINE**  
**PUBMED** 21819461  
**COMMENT** 11830664  
 Contact: Muller Miklos  
 Laboratory of Biochemical Parasitology  
 The Rockefeller University  
 1230 York Avenue, New York, NY 10021, USA  
 Email: mmuller@rockefeller.edu  
 Insert Length: 1550 Std Error: 0.00  
**POLYA=No.** Location/Qualifiers  
 1..1550  
 /organism="Mastigamoeba balamuthi"  
 /mol\_type="mRNA"  
 /strain="ATCC 30984"  
 /db\_xref="taxon:108607"  
 /clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
 /note="syn: Phreatamoeba balamuthi"

**FEATURES**  
 source  
 Query Match 3.1%; Score 76.6; DB 12; Length 1550;  
 Best Local Similarity 45.6%; Pred. No. 8.9;  
 Matches 308; Conservative 0; Mismatches 364; Indels 3; Gaps 1;

**ORIGIN**  
 1755 CGACGTGAGCAGCTGACCGAGCGCGTGCAGAAAGATCGCATGGAGAGCATCGTATCTG 1814  
 Db 5 CGGACGAGAGCGAGGACCGCGCGCGCGCATCGGAGCGCGGCGGCTCGTCTT 64  
 1815 GGCGAAGACCCCAAGTTCGGCTCCCATCCAGAGAGACCTGGGAGACCTGGTGCAG 1874  
 Db 65 CGCCTGGAAGGCGAGAACTCCAGAGTACTTGGAGTGCACCTGGAAGGCGCTGTCTT 124  
 1875 CGACTACTGGCAGGCGACCTGATCCGAGTGGGAGTTCGTGAACACCCCGCCCTGCT 1934  
 Db 125 CGGCCCTTACCAGGCGCTCAGATATCTGTCACGACGCGGTGACGACCTCTGATGAT 184  
 1935 GAAGCTGTGTATACAGCTGGAGAGAGCCCATATCTGCGCGCGGAGACCTTCTAGTGA 1994  
 Db 185 CCACAGGGGTTTCGCGCGGAGACAAACCCCAAGCTGTGGAGAGACGACGAGGCGCTCGA 244  
 1995 CGCGCGCGCAACCGGAGACCAAGATCGGCAAGCGCGCTACGTGACCGCGGCGG 2054  
 Db 245 GAGGTGCGCTGCTCAACAACTGCTCAAGCAGGTTCAGAGAGAGCAGCCCGCTTCTG 304  
 2055 GCAGAAAGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTCGAGGCCATCCA 2114  
 Db 305 GCACAGATCTCTCCCGAGATCCGCGTGTGACGAGGAGACGACGACTGGGTGATGAG 364  
 2115 GTTGGCGCTGAGAGCAGCGGAGCGAGGTGACATCTGTCGCGAGCAGCAGTACGCCCT 2174  
 Db 365 GCTGTATCCAGCTCACCGGACCGGAGCTGTGTTCCCGCGCGTCAAGTCAACGACTC 424  
 2175 GGGCATCATCCAGCGCGCGCGCGAGAGCGAGCTGGTGAACACAGATCATCA 2234  
 Db 425 TG---TCACCAAGAGCAAGTTTGACAACTTACGCTGCGCGCACTCGCTCATCGACG 481  
 2235 GCAGCTGATCAAGAGAGAGGTGTACTGAGCTGGTGGCGCGCGCAAGGCGCATCG 2294  
 Db 482 CATCAAGCGCGCAGCAGCTGATGTCTGCGCGCAAGGTGCGCGTCTGCGGCGGTACGG 541  
 2295 CGGCAAGCAGCAGATCGACAAGCTGGTGAAGAGGCGCATCCGCAAGGTGTGTCTCTGGA 2354  
 Db 542 CGACGTGGCAAGGGCTGCGCCAGTCTGCTCGCGCGCGGCTGCGCGGTCTCATCGTAC 601

**FEATURES**  
 source  
 Query Match 3.2%; Score 79; DB 12; Length 867;  
 Best Local Similarity 45.7%; Pred. No. 4.1;  
 Matches 314; Conservative 0; Mismatches 370; Indels 3; Gaps 1;

**ORIGIN**  
 58 CAGCGCGCAGCTTCAAGGCGCCCAAGCGGATCATCAAGTCTTCAACTCGCGCAAGGAG 117  
 Db 131 CCGCACAGATGCGCGAGTGTCTCGCGTCTATCTCTGCTGCGCAACAGCTTGAAGTAC 190  
 118 GGCACATCGCCCGCAACTGCGCGCGCCCGCGCAAGAGGCTCTGGAAGTGGCGCAAG 177  
 Db 191 GCGTGACCGCGCGTGAAGTCACTCGATCTGATGAGCGGCTGATCAAGATCGACGGC 250  
 178 GAGGCGCACCATGATGAGAGCTGACCGAGCGCGCGCAAGCTTCTTCCGAGGAGCTG 237  
 Db 251 AAGGTCCGACCGCACAGCACTTCCCGCGGCTTCAAGACGTCTGTCGATCGACAAG 310  
 238 GCCTTCCCGCGGCGGAGTTCCTCCCGAGCGAGCAACCGCGCGCAAGCGCC 297  
 Db 311 ACCGAGAGCACTTCCGCTGCTCTACGACACCAAGGCGCTTCCAGGCGCAACCGCATC 370  
 298 ACCAGCCCGAGTGCAGGT---GGCGCGCACAAACCCCGCAGCGGCGCGCGAG 354  
 Db 371 AACTCGACCGAGGCGCAAGTTCAAGTCCGCAAGTTCGCGCGCTGCAAGTCTGGCAACAG 430  
 355 GCGCAGGCGCACCTGAACTTCCCGCAGATCACTCTGTGGCAGCGCCCGCTTGTGAGCATC 414  
 Db 431 GGCATCCGTACTGTGACCCACGACGCGCGCAGTCCGCTACCCCAACCCCGCATC 490  
 415 AAGTGGCGGCGCAGATCAAGAGGCGCTGTGACACCGCGCGCGGCGAGCACAGTCTG 474  
 Db 491 AAGTCAACGACACAGTCAAGATCACTGCGCTGCGGCAAGATCATGATCTTCTGTCAGG 550  
 475 GAGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTC 534  
 Db 551 TTCGAGATCGGCAACTCTGTGATGATCACTGGCGGACGCAACCTTGGCGCGTCTCGGCTC 610  
 535 ATCAAGGTGCGCAGTACAGCAGATCTGTGATCGAGATCTGGGCAAGAGGCCATCGGC 594  
 Db 611 ATTGTGCGCGGAGAGCACGAGGCTGCTTCGAGATCATCCAGTCAAGGACGCGCTC 670  
 595 ACCGTGCTGATCGCGCCACCGCTGAAATCATCGCGCGCAACATGCTGACCCAGCTG 654  
 Db 671 GCGCACAGTTCGCGACGCGCTGACCAAGTCTTCTGATCGGCAAGGCGCAAGTCC 730  
 655 GGTGACACCTGAATTCCTCCATCAGCCCGATCGAGACCGTGCCTGGAAGTGAAGCCC 714  
 Db 731 CTGCTCAGCTGCGCGCGCGGAGGCGCATCAAGAAAGTCTGATCATCGAGGAGTTCCAGCG 790  
 715 GCATGAGCGCGCGCAAGGTGAAGCAG 741  
 Db 791 CGCCACGCGCACAGGACCGAGCAGGAG 817

**RESULT 7**  
 BM321022 1550 bp mRNA linear EST 03-JAN-2002  
 LOCUS  
 DEFINITION  
 rockefeller.0.1192 Mastigamoeba balamuthi lambda ZAP II Library  
 Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinease (EC  
 3.3.1.1), mRNA sequence.  
 ACCESSION  
 VERSION  
 BM321022.1 GI:18055428  
 KEYWORDS  
 EST.

QY 2355 CGGCATCGATGGCGGATCGTGTATCTACAGTACATGGACGACCTGTACGTGGGCGCG 2414  
 Db 602 GGAGATCGACCCCATCTGCGCGCTGCAGCGTGCATGGCGGGCTTCGAGGTCAACACGCT 661  
 QY 2415 CGGCCCTAGGATCGA 2429  
 Db 662 CGAGGCGGGGCTCGA 676

RESULT 8  
 BM321023 862 bp mRNA linear EST 03-JAN-2002  
 LOCUS rockefeller.0.594 Mastigamoeba balamuthi lambda ZAP II Library  
 DEFINITION Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (BC 3.3.1.1), mRNA sequence.  
 ACCESSION BM321023.1 GI:18055429  
 VERSION BM219461  
 KEYWORDS EST.  
 SOURCE Mastigamoeba balamuthi  
 ORGANISM Mastigamoeba balamuthi  
 Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.  
 REFERENCE 1 (bases 1 to 862)  
 AUTHORS Baptiste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W., Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and Philippe,H.  
 TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
 MEDLINE 21819461  
 PUBMED 11830664  
 COMMENT Contact: Muller Miklos  
 Laboratory of Biochemical Parasitology  
 The Rockefeller University  
 1230 York Avenue, New York, NY 10021, USA  
 Email: mmuller@rockvax.rockefeller.edu  
 Insert Length: 862 Std Error: 0.00  
 POLYA-No. Location/Qualifiers  
 1..862  
 /organism="Mastigamoeba balamuthi"  
 /mol\_type="mRNA"  
 /strain="ArCC 30984"  
 /db\_xref="taxon:108607"  
 /clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
 /note="syn: Phreatamoeba balamuthi"

FEATURES  
 source  
 1..862  
 Query Match 3.0%; Score 75; DB 12; Length 862;  
 Best Local Similarity 46.2%; Pred. No. 12;  
 Matches 283; Conservative 0; Mismatches 327; Indels 3; Gaps 1;

QY 1728 CGCCAAAGATGGCGACCGCCACACCAACGACGTGAAGAGCTGACCGAGCGCGTGCAGAA 1787  
 Db 220 CGCCAAAGATCGGTGCTGCTGTCGCAACATCTTCTCGACGCGAGCCACCGCGCGCGC 279  
 QY 1788 GATGCCATGGAGGACATCGTATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCATCCA 1847  
 Db 280 CATCGCGAGCGCGGAGTCTCGGTTTCGCTTGGAGGGCGAGAACCTCCAGGAGTACTG 339  
 QY 1848 GAAGGAGACCTGGGAGACCTGGTGACCGACTACTGGCAGGCGCACCTGGATCCCGAGTG 1907  
 Db 340 GGAGTGCACTINGAAGGCCCTGTGCTTCGGCCCTTACCAGGACCTCAGANCATCGTGA 399  
 QY 1908 GAGTTCTGTGAACACCCCGCCCTCGTGAAGTGTGTTACAGCTGGAGAGGAGCCCAT 1967  
 Db 400 CGACGGCGGTGACGCGACTCTAATGATCCAAAGGGGTTCGCGCGGAGGACACCCCAA 459  
 QY 1968 CATCGCGCCGAGACCTTCTAGTGGAGCGGGCGCCCAACCGCGAGACCAAGATCGGCAA 2027  
 Db 460 GCTGCTGAGAGACGAGGAGGCTCGAGAGGTTCGCTTCCTCAACACGTGCTCAAGCA 519  
 QY 2028 GGCCGGCTACGTACCGACCGGGCGCGAGAGATCGTGAAGCTTCGACCGGAGACCA 2087

Db 520 GGTCCAGAGAGAGCAGCCCGCTTCTGGCAACAAGATCTCTCCCGAGATCCGCGGTCTCAG 579  
 QY 2088 CCAGAAGACCGAGTGCAGGCCATCCAGCTGGCCCTGCAGACAGCGGCGAGCGGTGAA 2147  
 Db 580 CGAGAGACGACGACTGGCGTGTAGAGCTGTACAGCTGCACCGCGACGCGAAGCTGCT 639  
 QY 2148 CATCGTGACCGACAGCCAGTACGCCCTGGGCATC---ATCCAGGCCAGCCCGACAGAG 2204  
 Db 640 GTTCCCGCGGTCAACGTCAACGACTCTGTCAACAGAGCAAGTTTGACACATCTACGG 699  
 QY 2205 CGAGAGCGAGTGTGAACGAGATCATCGAGCAGCTGATCAAGAAGAGAGGTGTACCT 2264  
 Db 700 CTGCGGCCATCTGCTCATTCAGCGCATCAAGCGGCGCACCGATGATCTCGCGCGCAA 759  
 QY 2285 GAGCTGGTGGTCCGCCCCACAGGCGATCGCGGCGACGAGCAGATCCACAAGCTGTGTAG 2324  
 Db 760 GGTGCGCGTGTCTCGCGGCTACGCGCAGCTGGGCGAAGGCTGGCGCGAGTCTGTGCGCGG 819  
 QY 2325 CAAGGCGATCCGC 2337  
 Db 820 CCAGGCTGCCGC 832

RESULT 9  
 BZ568300 914 bp DNA linear GSS 17-DEC-2002  
 LOCUS pacs2-164\_7463.y2 pacs2-164 Pseudomonas aeruginosa genomic clone  
 DEFINITION pacs2-164\_7463, genomic survey sequence.  
 ACCESSION BZ568300  
 VERSION BZ568300.1 GI:27201058  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 REFERENCE 1 (bases 1 to 914)  
 AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.  
 TITLE Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library  
 JOURNAL J. Bacteriol. (2002) In press  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2062216954  
 Email: craymond@u.washington.edu  
 Class: shotgun.

FEATURES  
 source  
 1..914  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="2-164"  
 /db\_xref="taxon:287"  
 /clone="pacs2-164\_7463"  
 /clone\_lib="pacs2-164"  
 /note="clinical isolate 2-164 Whole genomic shotgun library."

ORIGIN  
 Query Match 3.0%; Score 74.4; DB 28; Length 914;  
 Best Local Similarity 47.0%; Pred. No. 15;  
 Matches 265; Conservative 0; Mismatches 296; Indels 3; Gaps 1;  
 QY 652 CTGGGTGACCTGTGACTTCCCATCGACCCCATCGAGACGTGCGCGTGAAGTGAAG 711  
 Db 112 CTGCGCAACAGCAGGACCGCTGTATCTTACGAGCGCGCGCTCTGAGGTGAGGCC 171  
 QY 712 CCCGGCATGAGCGGCCCAAGGTGAAGTGTAGCTGAGTGTGCGCTGACCGAGAGATCAAGGCC 771  
 Db 172 GGCAGCTGGACAAACCGCCAGGCGACGCTCCAGGCCAGGCGACAAACGAGTGTGCTATC 231

QY 772 CTGACGCCATCTGCGAGGATGGAGAGAGGCGAAGATCACCAAGATCGGCCCGAG 831  
Db 232 GGTGGCGCGGTGACAAACAGGCGCGCGCTGGACAGCGCGCGCGTAACTAGACTG 291  
QY 832 AACCCCTACAACACCCCGCTGTTCGCATCAAGAAGAGGACAGCACCAAGTGGCGCAAG 891  
Db 292 CAGAGCGGCGAGCTCGACAAACGCGCGCGCTGCTCAACAGCGCGCAAGGTTGGGTG 351  
QY 892 CTGGTGGACTTCGCGAGCTGAACAGAGCCACCCAGAGCTTCTGGAGGTGCAGTGGGC 951  
Db 352 AAGCTGGTCAACCGGTGTTCGACAAAGCGCGCGCTCACCAGCGCGAGTGCCTGGNA 411  
QY 952 ATCCCCCACCCCGCGCTGAAAGAAAGAGAGCGTGAACCGTGTGACGTGGCGGAC 1011  
Db 412 ATCCGCGCGCGGCAAGCGGTGCGCAACACGAGCGGC---CATGTCTCGCGCGTGGCGGC 468  
QY 1012 GCTACTTACGCTGCGCGCTGGAGAGGACTTCGCGAAGTACACCGCTTACCTACCTCCC 1071  
Db 469 GACAACCGCATCTGACCGCTGACTTCGACAAACAGGCGCGCGCTTACGCGCAGCGC 528  
QY 1072 AGCATCAACAAAGAGAGCCCGGCTACCGTACAGTACAACTGTGTGCGCCAGGCGTGG 1131  
Db 529 CTGCTAGCTCGACGCGCAGCGCTTCTCAACAGGCGCGCGCGCGCGCGCGC 588  
QY 1132 AAGGCGAGCGCGAGCATCTTCAGAGGAGCATGACCAAGATCTGGAGCGCTTCCGCGCG 1191  
Db 589 AAGTCTCGCGCGCGCGCATCTTACGCTGCGCGCGCGCGCGCGCGCGCGCGCGC 648  
QY 1192 CGCAACCCCGAGATCGGTACTAC 1215  
Db 649 AGTTGGAAGCGAGCGAGCTGC 672

RESULT 10  
BM321393  
LOCUS  
DEFINITION  
Mastigamoeba balamuthi 853 bp mRNA linear EST 03-JAN-2002  
rockefeller.0.1222 Mastigamoeba balamuthi lambda ZAP II Library  
Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA  
sequence.  
ACCESSION  
BM321393  
VERSION  
BM321393.1 GI:18055799  
KEYWORDS  
EST.  
SOURCE  
Mastigamoeba balamuthi  
ORGANISM  
Mastigamoeba balamuthi  
REFERENCE  
1 (bases 1 to 853)  
Baptiste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,  
Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and  
Philippe,H.  
The analysis of 100 genes supports the grouping of three highly  
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba  
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
21819461  
11830664  
Contact: Muller Miklos  
Laboratory of Biochemical Parasitology  
The Rockefeller University  
1230 York Avenue, New York, NY 10021, USA  
Email: mmuller@rockefeller.edu  
Insert Length: 853 Std Error: 0.00  
POLYA-No. Location/Qualifiers  
1..853  
/organism="Mastigamoeba balamuthi"  
/mol\_type="mRNA"  
/strain="ATCC 30984"  
/db\_xref="taxon:108607"  
/clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
/note="syn: Phreatamoeba balamuthi"

ORIGIN  
Query Match 3.0%; Score 73.2; DB 12; Length 853;  
Best Local Similarity 46.9%; Pred. No. 20;

Matches 299; Conservative 0; Mismatches 329; Indels 9; Gaps 2;  
QY 388 CTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGGCGGCAGATCAAGAGGCGCTTGCTG 447  
Db 125 CTGCGCGCGGTGTGTGAAGAAGCTCAACCTCGATCTCAAGTACGAGGTTGCAAGAAG 184  
QY 448 GACACCGCGCGCGAGCACACCGTGTGTGGAGAGATGAGCTGCCCGGCAAGTGGAGGCC 507  
Db 185 GTCAACGCGCGAGGACTACAACTCGAGGAGCTCGACGCGCGCGCGCTTCAAGGCC 244  
QY 508 AAGATGATCGCGCGCATCGCGCGCTTATCAAGGTGGCGCAGTACGACAGATCTCTGATC 567  
Db 245 CTGCTCGACGTGGCGCTGTGTCGCACTCGACCGCGCGCGCGGTGTCGCGCGCTCAAG 304  
QY 568 GAGATCTCGCGCAAGAAAGGCCATCGGCACCGTGTGATCGGCGCGCGCGCGTGAATC 627  
Db 305 GGCATGTGCGACGCGCGCGTCAACGTCGCTCC---CACAGCGAGACCGCGCTTCGTCGGCTTC 361  
QY 628 ATCGCGCGCACATCTGTGACCGAGTGGGTGGCTGCACCTGAATCTCCCATCAGCCCATC 687  
Db 362 AACCGCGCACAGAGAGGAGCTCAACGCGCGGTCTCTCGCAAGTACATCTTCGCGCGCAC 421  
QY 688 GAGACCGTGTGCGGTGAAGCTGAAGCCCGCATGGAAGCGCGCGCGCGCGCTTCGACCGCGCTTC 747  
Db 422 GTGCGCGGTACATGAAGCTCTCAAGGAGCAGGAGCGCGCGCGCTTCGACCGCGCTTC 481  
QY 748 CTGACCGAGGAGAGATCAAGGCGCTGACCGCGCATCGGAGGAGATGGAAGAGGAGGCC 807  
Db 482 TCGCGCTACGCGCAAGAGGAGGTGTCAACGCGCAATGCTCGAAGATCTACACGAGGCC 541  
QY 808 AAGATCACCAAGATCGCGCGCGCGCGCGCTTCAACACCGCGCGGTTCGCGCATCAAGAAG 867  
Db 542 CA-----CAGCAGATCGCGCGCGCGCGCGCTTCTCGCGCGCGCGCTCGAAGGCC 595  
QY 868 AAGACAGACCAAGTGGCGCGAGCTGTGTGAGCTTCGCGAGCTGGAACAAGCGCACCCAG 927  
Db 596 GAGGCGCGCAAGCCCAAGACATGCGGCGCAAGCGCGAGGTGAGTACACGAGCGCGCAAGAC 655  
QY 928 GACTTCTGGAGGTGCGAGCTGGGCATCCCGCGCGCGCGCGCGCTTGAAGAGAGAGAGG 987  
Db 656 CGGCTCGCGCGAGAGAGGTTCGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 715  
QY 988 GTGACCGTGTGAGGTGGCGCGCGCTTCTTCTGAGG 1024  
Db 716 CCGGTGCGCGGTACACNGCATCTCCGCGGTGCGCGG 752

BM587428 566 bp mRNA linear EST 25-FEB-2002  
17000687321202 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone  
19600449696110 5', mRNA sequence.  
BM587428  
BM587428.1 GI:18883289  
EST.  
Anopheles gambiae (African malaria mosquito)  
SOURCES  
Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;  
Anopheles.  
REFERENCE  
1 (bases 1 to 566)  
Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L.,  
Charlab, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.  
Celera Genomics  
Unpublished (2002)  
CONTACT: Holt R.A.  
Celera Genomics  
45 W. Gude Dr., Rockville, MD 20850, USA  
Tel: 2404533151  
Fax: 2404534580  
Email: HoltR@celera.com  
Plate: NU01004AX row: G column: 08  
Seq primer: M13 Reverse.



FEATURES

source

Location/Qualifiers  
1. 566  
/organism="Anopheles gambiae"  
/mol\_type="mRNA"  
/strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"  
/db\_xref="taxon:7165"  
/clones="19600449696110"  
/dev\_stage="Adult"  
/lab\_host="DH10b"  
/clone\_lib="A.Gam.ad.cdna.blood1"  
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen 24 hours after human blood feeding. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)"

ORIGIN

Query Match 3.0%; Score 72.8; DB 12; Length 566;  
Best Local Similarity 48.3%; Pred. No. 21;  
Matches 203; Conservative 0; Mismatches 217; Indels 0; Gaps 0;  
QY 1950 GTTGAGAGAGAGCCCATCATCGGCGCGAGACCTTCTACGTGAGCGCGCGCCGCAACCG 2009  
DB 482 GTTCAGATCGAGCGCAACCGCATCTCGAGGTGTGCGCGGAGGACAGGCGACGGGCAA 423  
QY 2010 CGAGACCAAGATCGCAAGCGCGGTAGTGCACGACCGCGCGCGCGAGAGATCGTGA 2069  
DB 422 CGGGAAGATCGTATCACCAGACGACCAAGACCGCTGACCGCGGAGCATCGAGCG 363  
QY 2070 CTTGACCGAGACCAACCAAGACGAGGTGCGAGCGCATTCACGTGGCCCTGCGAGGA 2129  
DB 362 CATGATCAAGGATCGGAGCGGTTCGCGAGCAAGCAAGAGTGAAGGAGCGGTGGA 303  
QY 2130 CAGCGGACGAGGTGAACATGTCGACGACGACGACGACGACGACGACGACGACGACG 2189  
DB 302 GCGCGCAACGAGCTCGAGAGTACGCGTACAGCCCTGAAGACCAAGCTCAGTCTGAAGGA 243  
QY 2190 CCAGCGCGACAGAGAGGAGCGAGTGTGTGAACAGATCATCGAGCAGCTGATCAAGAA 2249  
DB 242 CAAGTGGCGGAGCGGTTCGAGCAGACGACGACGACGACGACGACGACGACGACG 183  
QY 2250 GAGAGAGGTGTACTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2309  
DB 182 GAAGATCAAGTGGTGGAGAGAAACGAGGACCGGCGGAGAGTACAAAGACAGAA 123  
QY 2310 CGACAGCTGGTGAAGAGGATCGGCAAGGTGTCTTCTGGACGGCATCGATGGCGG 2369  
DB 122 GAAGGAGCTGGAGGACATCTGCGAGCCCATTCATTCGCAAGCTGTACGCGAGCAGTGGCGG 63

RESULT 12

CB643171

LOCUS  
DEFINITION CB643171 788 bp mRNA linear EST 08-APR-2003  
OSUNEB03L13.f OSUNEB Oryza sativa (japonica cultivar-group) cDNA  
clone OSUNEB03L13 5', mRNA sequence.

ACCESSION

VERSION

CB643171

KEYWORDS

EST.

SOURCE

ORGANISM

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 788)

Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,

Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.

Large-scale identification of ESTs involved in the interaction

between rice and Magnaporthe grisea

Unpublished (2003)

Contact: Rod Wing

Arizona Genomics Institute

University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: http://genome.arizona.edu  
PCR Primers  
FORWARD: gta aaa cga cgg cca gtc  
BACKWARD: gga aac agc tat gac cat g  
Plate: 03 row: L column: 13  
Seq primer: gta aaa cga cgg cca gtcg.

FEATURES

source

1. 788  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="OSUNEB03L13"  
/tissue\_type="Leaf"  
/dev\_stage="3 week"  
/lab\_host="DH10b"  
/clone\_lib="OSUNEB"  
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (Che 86061)"

ORIGIN

Query Match 2.9%; Score 72.6; DB 14; Length 788;  
Best Local Similarity 46.3%; Pred. No. 24;  
Matches 314; Conservative 0; Mismatches 355; Indels 9; Gaps 2;  
QY 107 GCGGCAAGAGGGCCCATCTGCGCGCACTGCGCGCCCCCGCAAGAGGCTGCTGA 166  
DB 99 GCGCGCGCGCGGAGATGGCGGCGAGCCCTCGACGAGGTGAAGCCATGTGGCGC 158  
QY 167 APTGCGCGAGGAGGCGCCACCATGACGAGTGCACCGAGCGGCCAGGCGCACTTTCC 226  
DB 159 AGTTCCGCGAGCGCTGCTCAAGATCCAGGGCGCCACCTCCGCGTCGCGCAGTGGCG 218  
QY 227 GCAGAGACTTGGCTTTCCCGAGGCAAGGCGCGAGTTCCTCCAGCAGCAGACCGCG 286  
DB 219 CCGTCCGCGAGGCAAGAGCGCGCGCGGTGCGCGTTCGAGTTCGACGAGGAGCGCGCC 278  
QY 287 CCACAGCCCCCAGCGCGGAGCTCAGGTGCGCGCGGACACACCCCGCAGCGAGCGCG 346  
DB 279 CCGCGCTCAAGGCGCAGCGAGTGTATCTCACTGCAATCGCCCAACGCGGCGCATCT 338  
QY 347 GCGCGAGCGCCAGGCGCACCTGAACTTCCCGCAGATCACCTGTGCGCAGCGCCCTGG 406  
DB 339 ACGGCTCACCA CGGCTTCGCGCGCACCTCCACCGCGCACCAAGGAGCGCGCGCC 398  
QY 407 TGAGCATCAAGGTGGCGGCGCATCAAGAGGCGCTGCTGACACCGGCGCGA---CG 463  
DB 399 TCCAAAGTCGAGCTCCTCAGGCATCTCAAGCGCGGAATCTTCGGCATTCGGTCCGATGGCC 458  
QY 464 ACACCGTGTGGAGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGGCGGCA 523  
DB 459 ACACGCTCGCTGGAGACGCTGCGGGCGCATGCTCGTGGCGCATCAACACCTCTCTCC 518  
QY 524 TCGCGGCTTCATCAAGGTGGCGGATGACGACGAGTTCCTGATTCGAGATTCGCGGCA 593  
DB 519 AGGCTACTTCGGCATCCGGTTCGAGATCTTCGAGGCGCATCAACAGCTGCTCAACACCG 578  
QY 584 AGGCTATCGGCA CGTGTCTGATCGCGCCACCCCGTGAACATCATCGGCGCGCAACATGC 643  
DB 579 GCGTCACCGCTGCTCGCGCTCCGTGGGACCATCATCCGCGTCCGCTGCTGCTGCTCC 638  
QY 644 TGACCCAGCTGGGTGACCCCTGAACCTTCCCATATGAGCCCATCGAGACCGTCCCGGTGA 703  
DB 639 TGTCTCATATTCGCGCTCATCACCGCGCGCCCAACGCGGAGGCGCATCTCGCGCGACG 698  
QY 704 AGCTGAAGCCCGGATGAGCGCGCCCAAGTGAAGTGGCCCTGACCGGAGGAGA 763

340 GCCTTCGGCGCACCGTCCACGTTGCAGACGCTGGCGCTCGAGCCCGCGGCGCTCGCGGACGAG 399

391 TGGCAGCGCCCCCTGTGTAGCATCAAGTGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450

400 GTCCGCGCGCACCTGCTCGGCTTGCCGCGCGCGCGGACCACCTACGCGCGGTGCGCCGC 459

451 ACCGCGCGCGACGACACACCGTGTGAGAGAGATGACCTGCCCGGAAAGTGAAGCCCAGAAG 510

460 GCTTGAAGCGGGGTACTGCTCCACG-----GCCGCGCGGCACGCGGCAAGACCAAGC 513

511 ATGATCGGGGATCGCGGGTTCATCAAGTGGCCAGTAGTACGACCAAGATCTGTATCGAG 570

514 CTCGTGCGTGCATCGCCAACTGCTGCAGCTTCGAGCTTACGACCTGAGGCTCACCAG 573

571 ATTCTCGGCAGAAAGGCCATCGGCACCGTGTGATCGGCCCGCCACCCCCGTTGAACATCATC 630

574 GTGCGCACCAACTCCCACTCGCGCGCTGCTCGTCTCCACGACCGCCCAAGTCCGCTCGTC 633

631 GGCGCCAA 638

634 GTCGTGCA 641

RESULT 14  
BE001575  
LOCUS  
DEFINITION  
HVSMH0098K08f Hordeum vulgare 5-45 DAP spike EST library  
HVCNDA0009 (S to 45 DAP) Hordeum vulgare subsp. vulgare cDNA clone  
HVSMH0098K08f, mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Hordeum vulgare subsp. vulgare  
Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.  
REFERENCE  
1 (bases 1 to 640)  
Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D.,  
Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W.,  
Fenton,R.D., Close,S.J., Oates,R. and Main,D.  
Development of a genetically and physically anchored EST resource  
for barley genomics: Morex 5-45 DAP spike cDNA library  
Unpublished (2001).  
JOURNAL  
COMMENT : On Aug 21, 2000 this sequence version replaced gi:13190104.

CONTACT: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: twing@clemson.edu  
Total hq bases = 461  
Seq primer: AATTACCTCTACTTAAGG3  
High quality sequence stop: 534.

FEATURES  
source  
Location/Qualifiers  
1..640  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Morex"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="HVSMH0098K08f"  
/tissue\_type="5-45 DAP Spike"  
/lab\_host="SOLR"  
/clone\_lib="Hordeum vulgare 5-45 DAP spike EST library"  
HVCNDA0009 (5 to 45 DAP)"  
/note=Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;  
plants were grown in the greenhouse at the University of  
California, Riverside (Fenton, SJ Close, TJ Close). Whole  
spikes with awns trimmed were collected at 5, 10, 15, 20,  
30 and 45 DAP (Fenton). total RNA was prepared from each  
pool, equal quantities of all six RNA pools were combined,

poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Choi) in the TJ Close lab at the University of California, Riverside. Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)

ORIGIN

Query Match 2.9%; Score 72.2; DB 10; Length 640;  
 Best Local Similarity 47.7%; Pred. No. 25;  
 Matches 209; Conservative 0; Mismatches 229; Indels 0; Gaps 0;  
 QY 789 GGAGATGAGAGAGAGAGATCACCAGATCGGCCCGAGAACCCCTACACACCCC 848  
 DB 110 GGAGAGGTGACGGTGAAGGTGTGGCCCAAGATGATCTCGGTGACGGGGCGCGGCAC 169  
 QY 849 CGTGTTCGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 908  
 DB 170 CCGTACCCGCACTTCAAGACACCTCACTCGACTTCCAGTTCGAGGACGCGGGCGCAA 229  
 QY 909 GTGTAACAG 968  
 DB 230 GCTCAAGTGGAGCGCTGTTCCGACCCGCGACCATGCGCCCATCGCACCGGCAT 289  
 QY 969 CCGTAAG 1028  
 DB 290 CTCACACGCTCCAGAACCTCATACCGCGGTCAACAAAGGGCTTCGCTACAAAGTGC 349  
 QY 1029 CCGTACAG 1088  
 DB 350 GGTCTACGCTACCTTCCCATCAAGCTTCATACCGCGCGCAACCGGGCATCGAGAT 409  
 QY 1089 CCGCGGATTCGCTACAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1148  
 DB 410 CGGCAACTTCTTCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 469  
 QY 1149 CTTCCAGAGCAGATGACCAAGATCTCGAGCCCTTCGCGCCCGCAACCCCGAGATCGT 1208  
 DB 470 CTTGCGGTCCGAGAGAGTCAAGATGATGATGATGATGATGATGATGATGATGATG 529  
 QY 1209 GATCTACAGAGCCCGCCCT 1226  
 DB 530 NTTCCGCTCGCGGCCCT 547

RESULT 15  
 CB648640  
 LOCUS  
 DEFINITION OSJNEB12C01.f OSJNEB Oryza sativa (japonica cultivar-group) cDNA clone OSJNEB12C01 5', mRNA sequence.  
 CB648640  
 ACCESSION  
 VERSION CB648640.1 GI:29643633  
 KEYWORDS EST.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzoideae; Oryza.  
 REFERENCE 1 (bases 1 to 688)  
 AUTHORS Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,

TITLE Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
 JOURNAL Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea  
 COMMENT Unpublished (2003)  
 Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA  
 Tel: 520 826 3967  
 Fax: 520 621 9288  
 Email: <http://genome.arizona.edu>  
 PCR Primers  
 FORWARD: gta aaa cga cgg cca gtg  
 BACKWARD: gga aac agc tat gac cat g  
 Plate: 12 row: C column: 01  
 Seq primer: gta aaa cga cgg cca gtg.  
 Location/Qualifiers

FEATURES

source

1..688  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJNEB12C01"  
 /tissue\_type="leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNEB"  
 /note="Vector: pBluescript II KS +; Site.1: EcoRI; Site.2: XhoI; 24 hrs after inoculation with Rice Blast (Che 86061)."  
 ORIGIN

Query Match 2.9%; Score 72.2; DB 14; Length 688;  
 Best Local Similarity 46.2%; Pred. No. 26;  
 Matches 276; Conservative 0; Mismatches 318; Indels 3; Gaps 1;  
 QY 107 GGGCAAG 166  
 DB 92 GCGCGCGCGCGCGAGATGGCGGAGCCACCTGACAGAGTGAAGCGCATGTGGCGC 151  
 QY 167 AGTGGCGCAAG 226  
 DB 152 AGTTCGCGAGCGGTGTCAAGATCCAGGGCGCCACCTCCGCGTCCGCCAGGTGGCGC 211  
 QY 227 GCGAG 286  
 DB 212 CCGTCGCCAGGCAAG 271  
 QY 287 CCAACAGCGCCACAGCGCGAGAGTCAAGTGGCGGCGACCAACCCCGCAGCGAGCGC 346  
 DB 272 CCGCGTCAAGCGCACAG 331  
 QY 347 GCGCGAG 406  
 DB 332 ACGGCGTCAACACCGGCTTCGCGGCGCACCTCCACCGCGCACCAAGGACGCGCGCGCC 391  
 QY 407 TGAGCATCAAGTGGCGGCGAGATCAAGAGAGCGCTGTGGACACCGCGCGCGA---CG 463  
 DB 392 TCCAAAGTCAGAGTCTCTCAGGATCTCAAGCGCGAGATCTTCGGCATGTGTCGATGGCC 451  
 QY 464 ACACCGTGTGGAG 523  
 DB 452 ACAGCTCGCGTCGGAG 511  
 QY 524 TCGGCGGCTTCATCAAGTGGCGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 583  
 DB 512 AGGGCTACTCCGGCATCCGGTTCGAGATCTTCGAGGCCATCACCAGCTGTCTCAACACCG 571  
 QY 584 AGGCGCATCGGACCGGTGTGATCGGCGCCACACCGCGCGTGAACATCATCGGCGCAACATGC 643  
 DB 572 GCGTCAGCGCGCTGCGCTCGCGCTCCGTGGGACCATCATCCGCGTCCGCTGACCTGTTCC 631



score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 02:53:16 ; Search time 6326.67 Seconds  
(without alignments)  
16873.640 Million cell updates/sec

Title: US-09-610-313-31  
Perfect score: 2463  
Sequence: 1 gtcagccaccatgcccga.....gggctagaccggtagaattc 2463

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Result No.	Score	Query Match	Length	DB ID	Description
1	2463	100.0	2463	6	AX455915 Sequence
2	2442.2	99.2	2469	6	AX455914 Sequence
3	2436.2	98.9	2457	6	AX455916 Sequence
4	2046	83.1	2306	6	BD263704 Improved
5	2027.6	82.3	2306	6	AR373387 Sequence
6	2025.2	82.2	2312	6	AX427930 Sequence
7	2025.2	82.2	2312	6	BD263706 Improved
8	2025.2	82.2	2312	6	AR373389 Sequence
9	2019.2	82.0	2300	6	BD263705 Improved
10	2019.2	82.0	2300	6	AR373388 Sequence
11	2019.2	82.0	2300	6	AX427936 Sequence
12	2005.2	81.4	2169	6	AX427931 Sequence
13	2001.8	81.3	2194	6	AX427926 Sequence
14	2000.2	81.2	2194	6	AX427925 Sequence
15	2000.2	81.2	2194	6	AX427927 Sequence
16	1996.8	81.1	2194	6	AX427938 Sequence
17	1993.4	80.9	2170	6	AX427933 Sequence
18	1993.4	80.9	2170	6	AX427928 Sequence
19	1993.4	80.9	2170	6	AX427935 Sequence
20	1993.4	80.9	2170	6	AX427932 Sequence
21	1993.4	80.9	2170	6	AX427934 Sequence
22	1991.8	80.9	2170	6	AX427921 Sequence
23	1988.8	80.7	2170	6	AF287353 Synthetic
24	1972.2	80.1	2170	6	AX455946 Sequence
25	1955.6	78.4	2170	6	AX455987 Sequence
26	1942.6	78.9	2170	6	BD263639 Improved
27	1942.6	78.9	2170	6	AR373322 Sequence
28	1932	78.4	2170	6	AF287352 Synthetic
29	1886.6	76.6	2170	6	AX457088 Sequence
30	1879	76.3	2170	6	AX306428 Sequence
31	1878.8	76.3	2170	6	BD263702 Improved
32	1878.8	76.3	2170	6	AR373385 Sequence
33	1875	76.1	2170	6	AX306429 Sequence
34	1865.4	75.7	2170	6	AX455954 Sequence
35	1857.8	75.4	2170	6	AX455952 Sequence
36	1852.4	75.2	2170	6	BD263703 Improved
37	1852	75.2	2170	6	AR373386 Sequence
38	1852	75.2	2170	6	AF287355 Synthetic
39	1798.8	73.0	2170	6	AX427937 Sequence
40	1784	72.4	2170	6	AR302563 Sequence
41	1651.8	67.1	2170	6	AX019132 Sequence
42	1651.8	67.1	2170	6	BD130468 Antiviral
43	1651.8	67.1	2170	6	BD268901 Anti-vira
44	1632.6	66.3	2170	6	AX035453 Sequence
45	1632.6	66.3	2170	6	AX035453 Sequence

ALIGNMENTS

RESULT 1	AX455915	AX455915	2463 bp	DNA	linear	PAT 06-JUL-2002
AX455915	LOCUS	Sequence 31 from Patent WO0204493.				
AX455915	DEFINITION	Sequence 31 from Patent WO0204493.				
AX455915	ACCESSION	AX455915				
AX455915	VERSION	AX455915.1	GI:21714900			
AX455915	KEYWORDS	synthetic construct				
AX455915	SOURCE	synthetic construct				
AX455915	ORGANISM	artificial sequences.				
AX455915	REFERENCE	zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.				
AX455915	AUTHORS	Polynucleotides encoding antigenic hiv type c polypeptides,				
AX455915	TITLE	polypeptides and uses thereof				
AX455915	JOURNAL	Patent: WO 0204493-A 31 17-JAN-2002;				

Pred. No. is the number of results predicted by chance to have a

CHIRON CORPORATION (US) : University of Stellenbosch (ZA)  
 Location/Qualifiers  
 1. 2463  
 /organism="synthetic construct"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32630"  
 /note="PR975YM"

ORIGIN

Query Match 100.0%; Score 2463; DB 6; Length 2463;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-249; Indels 0; Gaps 0;  
 Matches 2463; Conservative 0; Mismatches 0;

QY 1 GTCCAGCCGACACATGCGCCAGGCGCATGAGCCAGGCCACCCAGGCGCCAAATCTGTATGCGAG 60  
 Db 1 GTCCAGCCGACACATGCGCCAGGCGCATGAGCCAGGCCACCCAGGCGCCAAATCTGTATGCGAG 60

QY 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTCAACTGCGCGCAAGGAGGC 120  
 Db 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTCAACTGCGCGCAAGGAGGC 120

QY 121 CACATCGCCGCAACTGCGCGCCCGCCCGCAAGAGGGTGTGGAAGTGCAGGCAAGGAG 180  
 Db 121 CACATCGCCGCAACTGCGCGCCCGCCCGCAAGAGGGTGTGGAAGTGCAGGCAAGGAG 180

QY 181 GGCACACAGATGAGGACTGCACCGAGCGCGCGCCCAACTTCTTCGCGAGAGACTGGCC 240  
 Db 181 GGCACACAGATGAGGACTGCACCGAGCGCGCGCCCAACTTCTTCGCGAGAGACTGGCC 240

QY 241 TTCCCGCAGGAGGCGCCGCGAGTTCCCGCAGCGAGCAGAACCGCGCCCAACAGCCCCACC 300  
 Db 241 TTCCCGCAGGAGGCGCCGCGAGTTCCCGCAGCGAGCAGAACCGCGCCCAACAGCCCCACC 300

QY 301 AGCCGCGAGCTGCAAGTGCAGCGCGCAACACCCCGCAGAGAGCGCGCGCGAGCGCCAG 360  
 Db 301 AGCCGCGAGCTGCAAGTGCAGCGCGCAACACCCCGCAGAGAGCGCGCGCGAGCGCCAG 360

QY 361 GGCACCTGTAACTTCCCGCAGATCACCTGTGCGCGCGCCCTGTGAGCANTCAAGGTG 420  
 Db 361 GGCACCTGTAACTTCCCGCAGATCACCTGTGCGCGCGCCCTGTGAGCANTCAAGGTG 420

QY 421 GCGCGCGAGATCAAGGAGGCGCTGTGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGAG 480  
 Db 421 GCGCGCGAGATCAAGGAGGCGCTGTGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGAG 480

QY 481 ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATGCGCGCGCATCGCGCGCTTCAATCAAG 540  
 Db 481 ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATGCGCGCGCATCGCGCGCTTCAATCAAG 540

QY 541 GTGCGCGAGTACGACACAGATCCTGATCGAGATCTGCGCGCAAGAGGCCATCGGCACCGTG 600  
 Db 541 GTGCGCGAGTACGACACAGATCCTGATCGAGATCTGCGCGCAAGAGGCCATCGGCACCGTG 600

QY 601 CTGATCGCGCCCGACCCCGGTGAACATCATGCGCGCGCAACATGCTGACCCAGCTGGGTGC 660  
 Db 601 CTGATCGCGCCCGACCCCGGTGAACATCATGCGCGCGCAACATGCTGACCCAGCTGGGTGC 660

QY 661 ACCCTGAATTTCCCATCAGCCCATCAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCATG 720  
 Db 661 ACCCTGAATTTCCCATCAGCCCATCAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCATG 720

QY 721 GACGCGCCCAAGGTGAAGTGAAGTGGCCCTGACCGAGAGAGATCAAGGCGCTGACCGCC 780  
 Db 721 GACGCGCCCAAGGTGAAGTGAAGTGGCCCTGACCGAGAGAGATCAAGGCGCTGACCGCC 780

QY 781 ATCTCGAGGAGATGAGAGGAGGCGCAAGATCACCAAGATCGCGCCCGAGAACCCCTAC 840  
 Db 781 ATCTCGAGGAGATGAGAGGAGGCGCAAGATCACCAAGATCGCGCCCGAGAACCCCTAC 840

QY 841 AACACCCCGGTGTTCCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
 Db 841 AACACCCCGGTGTTCCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900

QY 901 TTCCGCGAGCTGAACAAGCGCAACCCAGGACTTCTGGAGGTGCAGTGGGCTATCCCCAC 960  
 Db 901 TTCCGCGAGCTGAACAAGCGCAACCCAGGACTTCTGGAGGTGCAGTGGGCTATCCCCAC 960

QY 961 CCGCGCGCTGAGAAAG 1020  
 Db 961 CCGCGCGCTGAGAAAG 1020

QY 1021 AGCGTGCCTGAG 1080  
 Db 1021 AGCGTGCCTGAG 1080

QY 1081 AAG 1140  
 Db 1081 AAG 1140

QY 1141 CCGAGCATCTTCAG 1200  
 Db 1141 CCGAGCATCTTCAG 1200

QY 1201 GAGATCGTATCTACAGGCGCCCTGTAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
 Db 1201 GAGATCGTATCTACAGGCGCCCTGTAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260

QY 1261 CCGCCCAAGATCGAG 1320  
 Db 1261 CCGCCCAAGATCGAG 1320

QY 1321 AAG 1380  
 Db 1321 AAG 1380

QY 1381 TGGACCGTGCAGGCGCATCGAGCTGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 Db 1381 TGGACCGTGCAGGCGCATCGAGCTGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440

QY 1441 AAGCTGTGGCAAGAGTGAAGTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 Db 1441 AAGCTGTGGCAAGAGTGAAGTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500

QY 1501 CTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGACATCGTGGCGCGCTGACCGAGAG 1560  
 Db 1501 CTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGACATCGTGGCGCGCTGACCGAGAG 1560

QY 1561 GCGAGCTGAGAGCTGGCGAG 1620  
 Db 1561 GCGAGCTGAGAGCTGGCGAG 1620

QY 1621 TACGACCCCGAG 1680  
 Db 1621 TACGACCCCGAG 1680

QY 1681 TACGAGATCTACAG 1740  
 Db 1681 TACGAGATCTACAG 1740

QY 1741 ACCGCGCGAG 1800  
 Db 1741 ACCGCGCGAG 1800

QY 1801 AGCATCTGTATCTGGGCGAG 1860  
 Db 1801 AGCATCTGTATCTGGGCGAG 1860

QY 1861 GAGACCTGTGTGAG 1920  
 Db 1861 GAGACCTGTGTGAG 1920

QY 1921 ACCCGCGCGCTGTGAG 1980  
 Db 1921 ACCCGCGCGCTGTGAG 1980

QY 1981 ACCTTCTACGTGAG 2040

```

Db      1981  ||||| ACCTTTACGTGACGGCGCGCAACCGCGAGACCAAGATCGCAAGCGCGGCTACGTG 2040
Qy      2041  ACCGACCGGGCGCGGAGAGATCGTAGCGCTGACCGGAGACCAACACAGAGACCGAG 2100
Db      2041  ACCGACCGGGCGCGGAGAGATCGTAGCGCTGACCGGAGACCAACACAGAGACCGAG 2100
Qy      2101  CTGAGGCCATCCAGCTGGCCCTGCGAGGACAGCGGAGCGAGGTGAACATCGTAGCCGAC 2160
Db      2101  CTGAGGCCATCCAGCTGGCCCTGCGAGGACAGCGGAGCGAGGTGAACATCGTAGCCGAC 2160
Qy      2161  AGCCAGTACGCCCTGGGCATCATCCAGGCCCGACCGACCAAGAGCGAGAGCGAGCTGGTG 2220
Db      2161  AGCCAGTACGCCCTGGGCATCATCCAGGCCCGACCGACCAAGAGCGAGAGCGAGCTGGTG 2220
Qy      2221  AACAGATCATCGAGCGAGCTGATCAAGAGGAGAGGTGTACTGAGCTGGGTGCCCGCC 2280
Db      2221  AACAGATCATCGAGCGAGCTGATCAAGAGGAGAGGTGTACTGAGCTGGGTGCCCGCC 2280
Qy      2281  CACAAGGCGCATCGCGGCAACAGCAGATCGACAAGCTGGTAGCAAGGGCATCCGCAAG 2340
Db      2281  CACAAGGCGCATCGCGGCAACAGCAGATCGACAAGCTGGTAGCAAGGGCATCCGCAAG 2340
Qy      2341  GTGCTGTTCCTGAGCGGATCGATGCGGCGATCGTGTATCAACAGTACATGAGAGCCTG 2400
Db      2341  GTGCTGTTCCTGAGCGGATCGATGCGGCGATCGTGTATCAACAGTACATGAGAGCCTG 2400
Qy      2401  TAGTGGCGAGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2460
Db      2401  TAGTGGCGAGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2460
Qy      2461  TTC 2463
Db      2461  TTC 2463

RESULT 2
AX455914
LOCUS      AX455914          2469 bp      DNA      linear      PAT 06-JUL-2002
DEFINITION Sequence 30 from Patent WO0204493.
ACCESSION  AX455914
VERSION     AX455914.1  GI:21714899
KEYWORDS    synthetic construct
SOURCE      synthetic construct
ORGANISM    artificial sequences.

REFERENCE
1
AUTHORS     zur Megede, J., Barnett, S.M., Engelbrecht, S. and van Rensburg, E.
TITLE       Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
JOURNAL     Patent: WO 0204493-A 30 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES
            Location/Qualifiers
                1..2469
                    /organism="synthetic construct"
                    /mol_type="unassigned DNA"
                    /db_xref="taxon:32630"
                    /note="PR975(+)"

ORIGIN
Query Match      99.2%; Score 2442.2; DB 6; Length 2469;
Best Local Similarity 99.6%; Pred. No. 8.6e-247;
Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

Qy      1  GTCGAGGCCATCGCGGAGGCCATGAGCCAGGCCACAGGCCCAACATCCTGATGCAG 60
Db      1  GTCGAGGCCATCGCGGAGGCCATGAGCCAGGCCACAGGCCCAACATCCTGATGCAG 60
Qy      61  CGCAGCACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTCGGGCAAGAGGGC 120
Db      61  CGCAGCACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTCGGGCAAGAGGGC 120
Qy      121  CACATCGCCCGCACTGCGCGCCCGCCCGCAAGAGGGCTGCTGGAGTGGCGGCAAGGAG 180

```

```

Db      121  CACATCGCCCGCAACTCGCGCGCCCGCGCAAGAGGGCTGCTGGAAGTSCGGCAAGGAG 180
Qy      181  GGCACACAGATGAAGGACTGCAACGAGGCGCACGACCACTTCTTCGCGAGGACCTGGCC 240
Db      181  GGCACACAGATGAAGGACTGCAACGAGGCGCACGACCACTTCTTCGCGAGGACCTGGCC 240
Qy      241  TTCCGCCAGGCAAGGGCCCGAGTTCCTCCAGCGAGCAGAACCGCGCCAAACAGCCCCACC 300
Db      241  TTCCGCCAGGCAAGGGCCCGAGTTCCTCCAGCGAGCAGAACCGCGCCAAACAGCCCCACC 300
Qy      301  AGCGCGAGCTGCGAGGTGCGCGCGACAAACCCCGCAGCGAGCGCGCGCGAGCGCGCAG 360
Db      301  AGCGCGAGCTGCGAGGTGCGCGCGACAAACCCCGCAGCGAGCGCGCGCGAGCGCGCAG 360
Qy      361  GSCACCTTGAACTTCCGCCAGATCACCTGTGGCAGCGCCCTCTGTGTGAGCATCAAGGTG 420
Db      361  GSCACCTTGAACTTCCGCCAGATCACCTGTGGCAGCGCCCTCTGTGTGAGCATCAAGGTG 420
Qy      421  GCGGCCAGATCAAGAGAGGCGCTGTGGACACCGCGCGCGAGCACCGTGTGGAGGAG 480
Db      421  GCGGCCAGATCAAGAGAGGCGCTGTGGACACCGCGCGCGAGCACCGTGTGGAGGAG 480
Qy      481  ATGAGCCTGCGCGCAAGTGAAGGCCCAAGATGATCGCGCGCATCGCGCGGCTTCATCAAG 540
Db      481  ATGAGCCTGCGCGCAAGTGAAGGCCCAAGATGATCGCGCGCATCGCGCGGCTTCATCAAG 540
Qy      541  GTGCGCCAGTACGACCAAGATCCTGTGAGACACCGCGCGCGAGGCGCATCGGCAACGTTG 600
Db      541  GTGCGCCAGTACGACCAAGATCCTGTGAGACACCGCGCGCGAGGCGCATCGGCAACGTTG 600
Qy      601  CTGATCGCGCCACCGCGGTGAACATCATCGCGCGCACATCTGACCGCGCTGGCTGC 660
Db      601  CTGATCGCGCCACCGCGGTGAACATCATCGCGCGCACATCTGACCGCGCTGGCTGC 660
Qy      661  ACCCTGAACTTCCCATCAGCCCATCGAGACCGTGTCCCGTGAAGCTGAACCGCGCATG 720
Db      661  ACCCTGAACTTCCCATCAGCCCATCGAGACCGTGTCCCGTGAAGCTGAACCGCGCATG 720
Qy      721  GACGCGCCCAAGTGAAGAGTGGCGCTGACCGCGAGAGAGTCAAGCGCTGACCGCC 780
Db      721  GACGCGCCCAAGTGAAGAGTGGCGCTGACCGCGAGAGAGTCAAGCGCTGACCGCC 780
Qy      781  ATCTGCGAGAGATGAGAGAGGAGGAGGAGTCAACAGATCGCGCGCGAGAACCCCTAC 840
Db      781  ATCTGCGAGAGATGAGAGAGGAGGAGGAGTCAACAGATCGCGCGCGAGAACCCCTAC 840
Qy      841  AACACCCCGTGTTCGCCATCAAGAGAGGAGGAGGAGTCAAGTGGCGGCAAGTGGTGAC 900
Db      841  AACACCCCGTGTTCGCCATCAAGAGAGGAGGAGGAGTCAAGTGGCGGCAAGTGGTGAC 900
Qy      901  TTCGCGAGCTGAACAGCGCACCGAGGACTTCTGCGAGGTGACGCTGGGCAATCCCGAC 960
Db      901  TTCGCGAGCTGAACAGCGCACCGAGGACTTCTGCGAGGTGACGCTGGGCAATCCCGAC 960
Qy      961  CCGCGCGGCTGAAGAGAGAGGAGCGTGAACCGTGTGAGAGTGGGCGAGCGCTACTTTC 1020
Db      961  CCGCGCGGCTGAAGAGAGAGGAGCGTGAACCGTGTGAGAGTGGGCGAGCGCTACTTTC 1020
Qy      1021  AGCGTCCCGCTGAGCGAGGACTTCCGAGTACACCGCTTCACCATCCCGAGCATCAAC 1080
Db      1021  AGCGTCCCGCTGAGCGAGGACTTCCGAGTACACCGCTTCACCATCCCGAGCATCAAC 1080
Qy      1081  AACGAGACCCCGGCTACCGCTACCAAGTACAAAGTGTCTCCCGAGGGCTGGAAGGCGAG 1140
Db      1081  AACGAGACCCCGGCTACCGCTACCAAGTACAAAGTGTCTCCCGAGGGCTGGAAGGCGAG 1140
Qy      1141  CCAGCATCTTCCAGAGCGATGACCAAGATCTGGAGCGCTTCGCGCGCGGACCGCC 1200
Db      1141  CCAGCATCTTCCAGAGCGATGACCAAGATCTGGAGCGCTTCGCGCGCGGACCGCC 1200
Qy      1201  GAGATCGTGTATACCA-----GGCGCGCTGTGAGCGAGCGACCTGGAGATCGGC 1254

```



Db 1201 GAGATCGTGATCTACCAAGTACATGAGCAGACCTGTAGTGGGACGACCTGGAGATCGGC 1260  
Qy 1255 CAGACCCGCGCCAAAGATCGAGAGCTGCGCAGACACTGCTGCGCTGGGGCTTACACACC 1314  
Db 1261 CAGCACCGCGCCAAAGATCGAGAGCTGCGCAGACACTGCTGCGCTGGGGCTTACACACC 1320  
Qy 1315 CCGCACAAGAAGCACCAGAAAGAGCCGCCCTTCTGTGGATGGGTACGAGCTGCACCCC 1374  
Db 1321 CCGCACAAGAAGCACCAGAAAGAGCCGCCCTTCTGTGGATGGGTACGAGCTGCACCCC 1380  
Qy 1375 GACAAGTGAACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGTGGACCGTGAACAC 1434  
Db 1381 GACAAGTGAACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGTGGACCGTGAACAC 1440  
Qy 1435 ATCCAGAAGCTGTGGGCGCAAGCTGAACCTGGGCGACCCAGATCTACCCCGGCATCAAGGTG 1494  
Db 1441 ATCCAGAAGCTGTGGGCGCAAGCTGAACCTGGGCGACCCAGATCTACCCCGGCATCAAGGTG 1500  
Qy 1495 CGCCAGCTGTGAAGCTGTGCGCGCGCAGAGCCCTGACCGACATCGTGCCTCCCTGACC 1554  
Db 1501 CGCCAGCTGTGAAGCTGTGCGCGCGCAGAGCCCTGACCGACATCGTGCCTCCCTGACC 1560  
Qy 1555 GAGAGGCGAGCTGTGAGCTGGCGGAGAACCCGAGATCTGCGGAGCCCGTGCACCGC 1614  
Db 1561 GAGAGGCGAGCTGTGAGCTGGCGGAGAACCCGAGATCTGCGGAGCCCGTGCACCGC 1620  
Qy 1615 GTGTACTACGACCCCGACCAAGACCTGTGTGGCGGAGATCCAGAGCAGGCGCCACGACAG 1674  
Db 1621 GTGTACTACGACCCCGACCAAGACCTGTGTGGCGGAGATCCAGAGCAGGCGCCACGACAG 1680  
Qy 1675 TGGACCTACAGATCTACCAAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAG 1734  
Db 1681 TGGACCTACAGATCTACCAAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAG 1740  
Qy 1735 ATGCGCACCGCCCGACCAAGACCTGTGTGGCGGAGATCCAGAGCAGGCGCCACGACAGTGC 1794  
Db 1741 ATGCGCACCGCCCGACCAAGACCTGTGTGGCGGAGATCCAGAGCAGGCGCCACGACAGTGC 1800  
Qy 1795 ATGAGAGATCTGTGATCTGGGCGAGACCCCAAGCTTCGCTGCGCTGCCATCCAGAGAGG 1854  
Db 1801 ATGAGAGATCTGTGATCTGGGCGAGACCCCAAGCTTCGCTGCGCTGCCATCCAGAGAGG 1860  
Qy 1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCGCACCTTGGATCCCGAGTGGGAGTTC 1914  
Db 1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCGCACCTTGGATCCCGAGTGGGAGTTC 1920  
Qy 1915 GTGAAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGGAGAGAGGCCATCATCGGC 1974  
Db 1921 GTGAAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGGAGAGAGGCCATCATCGGC 1980  
Qy 1975 GCCGAGACCTTCTACGTGGACCGCGCCCAACCGGAGACCAAGATCGGCAAGGCGCGC 2034  
Db 1981 GCCGAGACCTTCTACGTGGACCGCGCCCAACCGGAGACCAAGATCGGCAAGGCGCGC 2040  
Qy 2035 TAGCTGACCGACCGGCGCGGAGAGATCTGTAGCTGTGACCGAGACCAACACGAGAG 2094  
Db 2041 TAGCTGACCGACCGGCGCGGAGAGATCTGTAGCTGTGACCGAGACCAACACGAGAG 2100  
Qy 2095 ACCGAGCTGACGGCCATCTACGTGGCCCTGACGACAGCGGCGAGGAGTGAACATCGTG 2154  
Db 2101 ACCGAGCTGACGGCCATCTACGTGGCCCTGACGACAGCGGCGAGGAGTGAACATCGTG 2160  
Qy 2155 ACCGAGCGAGTACGCCCTGGGATCATCCAGGCCCGACCGACCAAGAGCGAGCGAG 2214  
Db 2161 ACCGAGCGAGTACGCCCTGGGATCATCCAGGCCCGACCGACCAAGAGCGAGCGAG 2220  
Qy 2215 CTGCTGAACCAAGATCATCGACAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGTG 2274  
Db 2221 CTGCTGAACCAAGATCATCGACAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGTG 2280  
Qy 2275 CCGCCCGACAGGCGCATCGCGCGCAACGAGCAGATCGAAGCTGTGTGACAGAGGCGATC 2334  
Db 2281 CCGCCCGACAGGCGCATCGCGCGCAACGAGCAGATCGAAGCTGTGTGACAGAGGCGATC 2340

Qy 2335 CGCAAGGTGCTGTTCTTGACCGGCATCGATGCGGCATCGTGTATCTACAGTACATGAC 2394  
Db 2341 CGCAAGGTGCTGTTCTTGACCGGCATCGATGCGGCATCGTGTATCTACAGTACATGAC 2400  
Qy 2395 GACCTGTAGTGGGCGAGCGCGCTTAGATCGATTAAAGCTTCCCGGGGCTAGCACC 2454  
Db 2401 GACCTGTAGTGGGCGAGCGCGCTTAGATCGATTAAAGCTTCCCGGGGCTAGCACC 2460  
Qy 2455 GGTGAATTC 2463  
Db 2461 GGTGAATTC 2469

RESULT 3  
AX455916  
LOCUS  
DEFINITION Sequence 32 from Patent WO0204493.  
ACCESSION AX455916  
VERSION AX455916.1 GI:21714901  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
artificial sequences.

REFERENCE 1  
AUTHORS zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,  
polypeptides and uses thereof  
JOURNAL Patent: WO 0204493-A 32 17-JAN-2002;  
CHIRON CORPORATION (US); University of Stellenbosch (ZA)

FEATURES  
source  
1..2457  
Location/Qualifiers  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="PR975YMMW"

ORIGIN  
Query Match 98.9%; Score 2436.2; DB 6; Length 2457;  
Best Local Similarity 99.6%; Pred. No. 3.6e-246;  
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

Qy 1 GTCGACGCCACCATGCGCGGAGCGGCATGAGCCAGCCACCGCCAGCCATCTCTGATGCGAG 60  
Db 1 GTCGACGCCACCATGCGCGGAGCGGCATGAGCCAGCCACCGCCAGCCATCTCTGATGCGAG 60  
Qy 61 CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTCAACTCGCGCAAGGAGGC 120  
Db 61 CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTCAACTCGCGCAAGGAGGC 120  
Qy 121 CACATGCGCCGCAACTGCGCGGCGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAGGAG 180  
Db 121 CACATGCGCCGCAACTGCGCGGCGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAGGAG 180  
Qy 181 GGCCACAGATGAAGGACTGCACCGAGCGCCAGCCAACTTCTTCCGCGAGGACCTTGGCC 240  
Db 181 GGCCACAGATGAAGGACTGCACCGAGCGCCAGCCAACTTCTTCCGCGAGGACCTTGGCC 240  
Qy 241 TTCCCGCCAGGCAAGGCGCGGAGTTCCCGAGGAGAGAACCGCGCCACAGCCCCCACC 300  
Db 241 TTCCCGCCAGGCAAGGCGCGGAGTTCCCGAGGAGAGAACCGCGCCACAGCCCCCACC 300  
Qy 301 AGCGCGAGCTGCAGGCTGCGCGGCAAAJCCCGCGCAGCGAGCGCGCGCGCGAGCGCCAG 360  
Db 301 AGCGCGAGCTGCAGGCTGCGCGGCAAAJCCCGCGCAGCGAGCGCGCGCGCGAGCGCCAG 360  
Qy 361 GGCACCTTGAATTCGCCCAAGATCACCTGTGGAGCGGCCCTTGGTGTAGATCAAGGTG 420  
Db 361 GGCACCTTGAATTCGCCCAAGATCACCTGTGGAGCGGCCCTTGGTGTAGATCAAGGTG 420  
Qy 421 GCGCGCGAGATCAGGAGCGCCCTCTGGACACCGCGCGCGAGCACACCGTCTCGAGGAG 480  
Db 421 GCGCGCGAGATCAGGAGCGCCCTCTGTGACACCGCGCGCGAGCACACCGTCTCGAGGAG 480

QY 481 ATGAGCTGCCCCGCAAGTGAAGCCAAAGATGATCGGCGGATCGGGGGCTTCAATCAAG 540  
 Db 481 ATGAGCTGCCCCGCAAGTGAAGCCAAAGATGATCGGCGGATCGGGGGCTTCAATCAAG 540  
 QY 541 GTGCGCCAGTACGACAGATCTCTGTCGAGATCTCGGCAAGAGCCCATCGGCAACCGTG 600  
 Db 541 GTGCGCCAGTACGACAGATCTCTGTCGAGATCTCGGCAAGAGCCCATCGGCAACCGTG 600  
 QY 601 CTGATCGGCCCCCAGCCCGCTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC 660  
 Db 601 CTGATCGGCCCCCAGCCCGCTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC 660  
 QY 661 ACCCTGAATCTCCCATCAGCCCATCGAGACCGTGCCTGAGCTGAAGCCCGGCGATG 720  
 Db 661 ACCCTGAATCTCCCATCAGCCCATCGAGACCGTGCCTGAGCTGAAGCCCGGCGATG 720  
 QY 721 GAGCGCCCCAAGGTGAAGCATGCGCCCTGACCGAGGAGAAATCAAGGCCCTGACCCGCC 780  
 Db 721 GAGCGCCCCAAGGTGAAGCATGCGCCCTGACCGAGGAGAAATCAAGGCCCTGACCCGCC 780  
 QY 781 ATCTGGAGAGATGGAAGAGAGGGGCAAGATCAACCAAGATCGGCCCGGAGAACCCCTAC 840  
 Db 781 ATCTGGAGAGATGGAAGAGAGGGGCAAGATCAACCAAGATCGGCCCGGAGAACCCCTAC 840  
 QY 841 AACACCCCGCTGTTCCGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
 Db 841 AACACCCCGCTGTTCCGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
 QY 901 TTCGCGAGTGAACAAAGCGCACCCAGGATTTCTGGAGGTGAGTGGGATCCCGCAC 960  
 Db 901 TTCGCGAGTGAACAAAGCGCACCCAGGATTTCTGGAGGTGAGTGGGATCCCGCAC 960  
 QY 961 CCGCGCGCTGAG 1020  
 Db 961 CCGCGCGCTGAG 1020  
 QY 1021 AGCGTGCCCTGGACGAGGATTTCCGCAAGTACACCGCCCTTACCATCCCGACGATCAAC 1080  
 Db 1021 AGCGTGCCCTGGACGAGGATTTCCGCAAGTACACCGCCCTTACCATCCCGACGATCAAC 1080  
 QY 1081 AACGAGACCCCGGATCCCGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1140  
 Db 1081 AACGAGACCCCGGATCCCGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1140  
 QY 1141 CCCAGCATTTCCAGAGCAGCATGACCAAGATCTCGGAGCCCTTCCGCGCCCGCAACCCC 1200  
 Db 1141 CCCAGCATTTCCAGAGCAGCATGACCAAGATCTCGGAGCCCTTCCGCGCCCGCAACCCC 1200  
 QY 1201 GAGATCGGTGATCTACAGGCCCCCTGTGATCGTGGGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
 Db 1201 GAGATCGGTGATCTACAGGCCCCCTGTGATCGTGGGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
 QY 1261 CGGCGCAAGATCGAGGAGCTGCGAAGCACTGTGCGCTGGGCTTCAACACCCCGGAC 1320  
 Db 1261 CGGCGCAAGATCGAGGAGCTGCGAAGCACTGTGCGCTGGGCTTCAACACCCCGGAC 1320  
 QY 1321 AAGAAGCAACGAG 1380  
 Db 1321 AAGAAGCAACGAG 1380  
 QY 1381 TGGACGCTGAGCCATCGAGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 Db 1381 TGGACGCTGAGCCATCGAGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 QY 1441 AAGCTGTGGGCAAGCTGAATGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 Db 1441 AAGCTGTGGGCAAGCTGAATGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 QY 1501 CTGTGCAAGCTGCTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
 Db 1501 CTGTGCAAGCTGCTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
 QY 1495 CTGTGCAAGCTGCTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1554  
 Db 1495 CTGTGCAAGCTGCTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1554

QY 1561 GCCGAGCTGAGCTGGCCGAGAACCCGAGATCCTTCGCGAGCCCTGTGACCGCGGTGTAC 1620  
 Db 1555 GCCGAGCTGAGCTGGCCGAGAACCCGAGATCCTTCGCGAGCCCTGTGACCGCGGTGTAC 1614  
 QY 1621 TACGACCCCAAG 1680  
 Db 1615 TACGACCCCAAG 1674  
 QY 1681 TACGAGATCTACGAG 1740  
 Db 1675 TACGAGATCTACGAG 1734  
 QY 1741 ACCGCCCAACCAAG 1800  
 Db 1735 ACCGCCCAACCAAG 1794  
 QY 1801 AGCATCTGATCTGGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860  
 Db 1795 AGCATCTGATCTGGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1854  
 QY 1861 GAGACCTGTGTGAG 1920  
 Db 1855 GAGACCTGTGTGAG 1914  
 QY 1921 ACCGCCCGCTGTGAG 1980  
 Db 1915 ACCGCCCGCTGTGAG 1974  
 QY 1981 ACCCTTCTACGTGAG 2040  
 Db 1975 ACCCTTCTACGTGAG 2034  
 QY 2041 ACCGACCGGGCGGAG 2100  
 Db 2035 ACCGACCGGGCGGAG 2094  
 QY 2101 CTGCGAGGCTATCCAGCTGGCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160  
 Db 2095 CTGCGAGGCTATCCAGCTGGCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2154  
 QY 2161 AGCCAGTACGCGCTGGGAGATCATTCAGGCCCCAGGCGGAGAGAGAGAGAGAGAGAG 2220  
 Db 2155 AGCCAGTACGCGCTGGGAGATCATTCAGGCCCCAGGCGGAGAGAGAGAGAGAGAGAG 2214  
 QY 2221 AACCGAGATCTGAG 2280  
 Db 2215 AACCGAGATCTGAG 2274  
 QY 2281 CACAGGGGATCGGCGGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340  
 Db 2275 CACAGGGGATCGGCGGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2334  
 QY 2341 GTGCTGTTCTTGGAG 2400  
 Db 2335 GTGCTGTTCTTGGAG 2394  
 QY 2401 TACGTGGGAG 2460  
 Db 2395 TACGTGGGAG 2454  
 QY 2461 TTC 2463  
 Db 2455 TTC 2457

RESULT 4  
 BD263704  
 LOCUS  
 DEFINITION  
 Improved expression of HIV polypeptides and production of virus-like particles.  
 ACCESSION  
 BD263704  
 VERSION  
 BD263704.1  
 KEYWORDS  
 GI:33073472  
 JP 2002533124-A/71.  
 2306 bp DNA linear PAT 17-JUL-2003

SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE artificial sequences.  
1 (bases 1 to 2306)  
AUTHORS Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.  
TITLE Improved expression of HIV polypeptides and production of virus-like particles  
JOURNAL Patent: JP 2002533124-A 71 08-OCT-2002;  
CHIRON CORP  
COMMENT OS Artificial Sequence  
FN JP 2002533124-A/71  
PD 08-OCT-2002  
PR 30-DEC-1999 JP 2000591193  
PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 FI  
SUSAN BARNETT,JAN ZUR MEDEDE,INDRESH SRIVASTAVA,YING LIAN, PI  
KARIN HARTOG,  
DI HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER,PC  
C12N15/09,A61K31/71,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC  
C12N5/10,  
PC C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC  
Description of Artificial Sequence: FS(-).protmod.RTopt.YM FH Key  
Location/Qualifiers  
FT source 1..2306  
FT /organism='Artificial Sequence'.  
Location/Qualifiers  
1..2306  
/organism='synthetic construct'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32630'  
ORIGIN  
Query Match 83.1%; Score 2046; DB 6; Length 2306;  
Best Local Similarity 93.6%; Pred. No. 2.4e-205;  
Matches 2159; Conservative 0; Mismatches 135; Indels 12; Gaps 2;  
QY 170 GCGCAAGGAGGCGCCACAGATGAAGACTGCACCGAGCGCCAGGCGCAACTTTTCCGCG 229  
DB 1 GCGCGCGGAGGACCAATGAAGATTGCATGAGAGACAGGCTAATTTTCCGCG 60  
QY 230 AGGACTTGGCTTCCCGAGGCGAGCCCGCGAGTTCCCGAGGAGCGAGCGCGCA 289  
DB 61 AGGACTTGGCTTCCCGAGGCGAGCCCGCGAGTTTCAGCAGCGAGCAGACCCGCGCA 120  
QY 290 ACAGCCCCACAGCGCGAGCTGCAGGTGCGCGCGG-----ACAAACCCCGCAGCGAGG 343  
DB 121 ACAGCCCCACCGCGGAGCTGCAGGTGTGGGCGCGGAGACACAGCCTGAGCGAGG 180  
QY 344 CCGCGCGGAGCGCGAGGCGACCCCTG-----AACTTCCCCAGATCACTCTGTGCGAGC 397  
DB 181 CCGCGCGCGAGCGCGAGGCGACCGTGAGCTTCAACTTCCCCAGATCACTCTGTGCGAGC 240  
QY 398 GCCCTCTGTGAGCATCAAGTGGCGCGCGAGATCAAGGAGGCGCTGCTGGACACCGCGG 457  
DB 241 GCCCTCTGTGAGCATCAGATCGCGCGCGAGCTCAAGGAGGCGCTGCTGGACACCGCGG 300  
QY 458 CCGACGACACCGTCTGGAGGAGATGAGCTTGGCGCGAAGTGGAGCGCCCAAGATGTCG 517  
DB 301 CCGACGACACCGTCTGGAGGAGATGAACCTTGGCGCGAAGTGGAGCGCCCAAGATGTCG 360  
QY 518 CGCGCATCGCGGCTTCATCAAGTGGCGCGAGTACGACACAGATCTCTGATCGAGATCTCG 577  
DB 361 CGCGCATCGCGGCTTCATCAAGTGGCGCGAGTACGACACAGATCCCCGTGAGATCTGCG 420  
QY 578 GCAAGAGCGCATCGGCAACCGTGTGATCGGCCCGACCCCGTGAACATCATCGCGCGCA 637  
DB 421 GCCAAGAGCGCATCGGCAACCGTGTGATCGGCCCGACCCCGTGAACATCATCGCGCGCA 480  
QY 638 ACATGCTGACCGAGTGGCTGCACTTCCCTGACTTCCCATCAGCCCATCGAGACCTGCG 697  
DB 481 ACCTGCTGACCGAGATCGGCTGCACTTCCCTGACTTCCCATCAGCCCATCGAGACCTGCG 540  
QY 698 CCGTGAAGCTGAAGCGCGCGATGGACGCGCGCCCAAGGTGAAGAGTGGCGCCCTGACCGAGG 757

541 CCGTGAAGCTGAAGCGCGGATGGACGGCCCAAGGTCAAGCAGTGGCCCTGACCGAGG 600  
QY 758 AGAAGATCAAGGCCCTTGACCGGCATCTCGAGAGATGGAGAGAGGCGCAAGATACCA 817  
DB 601 AGAAGATCAAGGCCCTTGTTGGAGATCTGCACCGAGATGAGAAGGAGGCAAGATCAGCA 660  
QY 818 AGATCGGCCCGGAGAACCCCTTACACACACCCCGTGTTCGCCATCAAGAGAGGAGCAGCA 877  
DB 661 AGATCGGCCCGGAGAACCCCTTACACACACCCCGTGTTCGCCATCAAGAGAGGAGCAGCA 720  
QY 878 CCAAGTGGGCAAGCTGGTGGACTTCCGCGAGCTGAAACAGCGCACCCAGGACTTCTGGG 937  
DB 721 CCAAGTGGGCAAGCTGGTGGACTTCCGCGAGCTGAAACAGCGCACCCAGGACTTCTGGG 780  
QY 938 AGTGTGAGCTGGGCATCCCGCCCGCGGCTGAGAGAGAGAGAGCGGTGACCGTGC 997  
DB 781 AGTGTGAGCTGGGCATCCCGCCCGCGGCTGAGAGAGAGAGAGAGCGGTGACCGTGC 840  
QY 998 TGGACGTGGCGGACGCTACTTTCAGCGTGGCGCGGCTGAGAGAGAGAGAGAGCGGTGACCG 1057  
DB 841 TGGACGTGGCGGACGCTACTTTCAGCGTGGCGCGGCTGAGAGAGAGAGAGAGCGGTGACCG 900  
QY 1058 CTTTACCATCCCGCAGCATCAACAGAGACCCCGGATCCGCTACCGATCAAGTGC 1117  
DB 901 CTTTACCATCCCGCAGCATCAACAGAGACCCCGGATCCGCTACCGATCAAGTGC 960  
QY 1118 TGCCCCAGGGCTGGAAGGCGAGCCCGAGCATCTTCAGAGCAGCATGACCAAGATCCTGG 1177  
DB 961 TGCCCCAGGGCTGGAAGGCGAGCCCGGATCTTCAGAGCAGCATGACCAAGATCCTGG 1020  
QY 1178 AGCCCTTCCGCGCGCGCAACCCCGAGATCTGATCTACAGAGGCGCCCTTCTGCTGGGCA 1237  
DB 1021 AGCCCTTCCGCGCGAGCAACCCCGATCTGATCTACCAAGGCGCCCTTCTGCTGGGCA 1080  
QY 1238 CGGACCTGGAGATCGGCGAGCAGCGCGCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGC 1297  
DB 1081 CGGACCTGGAGATCGGCGAGCAGCGCGCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGC 1140  
QY 1298 GCTGGGGCTTCAACACCCCGGACAAAGAGCAACAGAGAGGAGCGCCCTTCTGCTGGATGG 1357  
DB 1141 GCTGGGGCTTCAACACCCCGGACAAAGAGCAACAGAGAGGAGCGCCCTTCTGCTGGATGG 1200  
QY 1358 GCTAGAGCTGCAACCGGACAAAGTGGAGCGGTGAGCCCATCGAGCTGCCGAGAGAGAGA 1417  
DB 1201 GCTAGAGCTGCAACCGGACAAAGTGGAGCGGTGAGCCCATCGAGCTGCCGAGAGAGAGA 1260  
QY 1418 GCTGACCGTGAAACGACATCCAGAGCTGGTGGGCAAGCTGAACTGGGCGCAGCGAGATCT 1477  
DB 1261 GCTGACCGTGAAACGACATCCAGAGCTGGTGGGCAAGCTGAACTGGGCGCAGCGAGATCT 1320  
QY 1478 ACCCGCGCATCAAGTGGCGCGAGCTGTGCAAGCTGTGCGGCGCGCCAGGCGCTGACCG 1537  
DB 1321 ACCCGCGCATCAAGTGGCGCGAGCTGTGCAAGCTGTGCGGCGCGCCAGGCGCTGACCG 1380  
QY 1538 ACATCGTCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGC 1597  
DB 1381 AGGTGATCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGA 1440  
QY 1598 GCGAGCGCGTGCACGCGGTGATCTAGACCCCGAGAGGAGCTGGTGGCGGAGATCCAGA 1657  
DB 1441 AGGAGCGCGTGCACGAGGTGATCTACGACCCCGAGGAGCTGGTGGCGGAGATCCAGA 1500  
QY 1658 AGCAGGCGCACGACGAGTGGACCTACAGATCTACAGAGCGCTTCAAGAACCTGAGA 1717  
DB 1501 AGCAGGCGCAGGCGAGTGGACCTTACAGATCTACAGAGCGCTTCAAGAACCTGAGA 1560  
QY 1718 CCGGCAAGTACGCCAAGATGCGCACCGGCCACACCAACGAGCTGAAGCAGCTGACCGAGG 1777  
DB 1561 CCGGCAAGTACGCCCGCATGGCGCGGCCACACCAACGAGCTGAAGCAGCTGACCGAGG 1620  
QY 1778 CCGTGCAGAGATCCGCTAGGAGCATCGTGTGGGCGAAGACCCCAAGTTCGCC 1837

1	CGCGCGCGAAGGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG	60
230	AGGACCTGGCTTCCCTCCAGGGCAAGCGCGAGTTCCCGAGCAGACAGACCCGCGCCA	289
61	AGGACCTGGCTTCTCTGAGGGCAAGCCCGCGAGTTCAAGCAGAGCAGACCCGCGCCA	120
290	ACAGCCCCACCGACCGCGAGCTCAGGTGCGCGCG-----ACAAACCCCGCAGCGAGG	343
121	ACAGCCCAACCCCGCGAGCTCAGGTGCGCGCGAGAAACAACAGCCTGAGGAGG	180
344	CCGCGCGCAGCGCAGGGCAACCTG-----AACTTCCCGCAGATCACCTCTGCAGC	397
181	CCGCGCGCAGCGCAGGGCAACCTGAGCTTCACTTCTCCCGCAGATCACTCTGCGCAGC	240
398	GCCTCTGTGTAGATCAAGTGGCGCGCAGATCAAGAGAGCCCTGTGTGAACACCGCG	457
241	GCCTCTGTGTAGATCAGGATCGCGCGCAGCTCAAGAGAGCGCTGCTCGACACCGCG	300
458	CCGACGACACCGTCTGTGAGAGATGAGCTGCCCGGCAAGTGAAGACCCCAAGATGATCG	517
301	CCGACGACACCGTCTGTGAGGAGATGAACCTGCCCGCAAGTGAAGACCCCAAGATGATCG	360
518	CGCGCATCGCGCTTATCAAGTGGCGCGCAGATCAAGACAGATCCTGTATCGAGATCTGCG	577
361	CGCGGATCGGGGCTTCAATCAAGTGGCGCAGTACACAGATATCCCGTGGAGATCTGCG	420
578	GCAAGAGGCATCGGCACCGTCTGATCGGCCCAACCCCGGTGAACATCATCGCGCGCA	637
421	GCCACAAGGCATCGGCACCGTCTGCTGGCGCCCAACCCCGGTGAACATCATCGCGCGCA	480
638	ACATGTGTACCGAGCTGGGCTGCACCTGAATTTCCCATCAGCCCATCGAGACCGTGC	697
481	ACTTGTGTACCGAGATCGGCTGCACCTGAATTTCCCATCAGCCCATCGAGACCGTGC	540
698	CCGTGAAGTGAAGCCCGCATGACCGGCCCAAGGTGAAGCAGTGGCCCTTGAACGAGG	757
541	CCGTGAAGTGAAGCCCGGATGAAGCGGCCCAAGGTCAAGCAGTGGCCCTTGAACGAGG	600
758	AGAGATCAAGGCCCTTGAACCGCATCTCGAGAGATGAGAGAGGGCCAGATCAACA	817
601	AGAGATCAAGGCCCTTGTGTGAGATCTGACCGAGATGAGAGAGGGCCAGATCAGCA	660
818	AGATCGGCCCGAGAACCCCTACAACACCCCGTGTTCGCCATCAAGAAAGAGCAGCA	877
661	AGATCGGCCCGAGAACCCCTACAACACCCCGTGTTCGCCATCAAGAAAGAGCAGCA	720
878	CCAAGTGGCGCAGCTGTGTGATCTCCGCGAGCTGAACAGCGCACCCAGGATCTCTGG	937
721	CCAAGTGGCGCAGCTGTGTGATCTCCGCGAGCTGAACAGCGCACCCAGGATCTCTGG	780
938	AGGTGAGCTGGGCATCCCGCCACCCCGCGGCTGAAGAAAGAGAGCGTGAACCGTGC	997
781	AGGTGAGCTGGGCATCCCGCCACCCCGCGGCTGAAGAAAGAGAGCGTGAACCGTGC	840
998	TGACGCTGGCGCAGCCTACTTCAAGTGCCTTGAACAGAGATCTCCGAGATCACCG	1057
841	TGACGCTGGCGCAGCCTACTTCAAGTGCCTTGAACAGAGATCTCCGAGATCACCG	900
1058	CTTTCACCATCCCGAGCATCAACAGAGACCCCGGCATCCGCTACAGTAAACGTGC	1117
901	CTTTCACCATCCCGAGCATCAACAGAGACCCCGGCATCCGCTACAGTAAACGTGC	960
1118	TGCCCCAGGCGCTGGAAGGGCAGCCCCAGCATCTTCAGAGCAGCATGACCAAGATCCTGG	1177
961	TGCCCCAGGCGCTGGAAGGGCAGCCCCAGCATCTTCAGAGCAGCATGACCAAGATCCTGG	1020
1178	AGCCCTTCGCGCCGCAACCCCGAGATCGTGATCTTACAGGCCCTCTGTACGTGGCA	1237
1021	AGCCCTTCGCGAAGCAGAACCCCGCATCTGTATCTACAGGCCCTCTGTACGTGGCA	1080
1238	GCAGCTTGAGATCGGCCAGCACCGCGCCCAAGATCGAGAGTGGCGAAGCATCTCTGC	1297



QY	542	TGCGCCAGTACGACCAGATCCTGTATCGAGATCTGCGGCAAGAGCCATCTGGCACCGTGC	601
Db	3500	TGGGCCAGTACGACCAGATCCTGTATCGAGATCTGCGGCCACAAGGCCATCGCACCCGTGC	3559
QY	602	TGATCGGCCCCACCCCGGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA	661
Db	3560	TGTTGGGCCCCACCCCGGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA	3619
QY	662	CCCTGAACTTCCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATGG	721
Db	3620	CCCTGAACTTCCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATGG	3679
QY	722	ACGGCCCAAGGTGAAGCAGTGGCCCTCAGCGAGGAGAGATCAAGGCCCTTGACCGCA	781
Db	3680	ACGGCCCAAGGTGAAGCAGTGGCCCTCAGCGAGGAGAGATCAAGGCCCTTGTTGGAGA	3739
QY	782	TCTTCGAGGAGATGGAGAAGAGGGCAAGATCACCAAGATCGGCCCGCGAAGACCCCTACA	841
Db	3740	TCTTCACCGAGATGGAGAAGAGGGCAAGATCAGCAAGATCGGCCCGCGAAGACCCCTACA	3799
QY	842	ACACCCCGCTGTTCGCCATCAAGAAAGGACACCAAGTGGCGCAGACTGGTGACT	901
Db	3800	ACACCCCGCTGTTCGCCATCAAGAAGAGGACACCAAGTGGCGCAGACTGGTGACT	3859
QY	902	TCCGCGAGCTGAAACAGCGCACCCAGACTTCTGGGAGGTGACGTGGGCATCCCCCACC	961
Db	3860	TCCGCGAGCTGAAACAGCGCACCCAGACTTCTGGGAGGTGACGTGGGCATCCCCCACC	3919
QY	962	CCGCGCGCTTGAAGAAAGAGGGTGACCGTGTGGAAGTGGGCGAGCGCTACTTCA	1021
Db	3920	CCGCGCGCTTGAAGAAAGAGGGTGACCGTGTGGAAGTGGGCGAGCGCTACTTCA	3979
QY	1022	GCGTGCCCTTGGACAGGACTTCCGCAAGTACACCGCTTACCATCCCCAGCATCAACA	1081
Db	3980	GCGTGCCCTTGGACAGGACTTCCGCAAGTACACCGCTTACCATCCCCAGCATCAACA	4039
QY	1082	ACGAGACCCCGGCATCCGCTPACAGTACAAACGTGCTGCCCAAGGTGGAAGGGCAGCC	1141
Db	4040	ACGAGACCCCGGCATCCGCTPACAGTACAAACGTGCTGCCCAAGGTGGAAGGGCAGCC	4099
QY	1142	CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCGCAACCCCG	1201
Db	4100	CCGCATCTTCCAGTGACAGTACCAAGATCTTGGAGCCCTTCCGCGCAAGCAACCCCG	4159
QY	1202	AGATCGTGATATACCA-----GGCCCGCCCTGTACGTGGGCAGCACTGGAGATCGGCC	1255
Db	4160	ACATCGTGATCTACCAGTACATGACACCTGTACGTGGGCAGCACTGGAGATCGGCC	4219
QY	1256	AGCACCGCCCAAGATCAGAGAGTGGCAAGCACCTGTGCGCTGGGGCTTCAACACCC	1315
Db	4220	AGCACCGCACCAAGATCAGAGAGTGGCCAGCACCTGTGCTGGGGCTTCAACACCC	4279
QY	1316	CCGCAAGAGCACCAAGAGAGCCCGCTTCTGTGGATGGGCTACGAGCTGCACCCCG	1375
Db	4280	CCGCAAGAGCACCAAGAGAGCCCGCTTCTGTGGATGGGCTACGAGCTGCACCCCG	4339
QY	1376	ACAAGTGAACCGTCAGCCCATCGAGCTGCCCGCAAGAGAGCTGACCGTGAACGACA	1435
Db	4340	ACAAGTGAACCGTCAGCCCATCGTGTCTGCCCGCAAGAGAGCAAGCTGGAGCGACA	4399
QY	1436	TCCAGAAGCTGGTGGCAAGCTGACCTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC	1495
Db	4400	TCCAGAAGCTGGTGGCAAGCTGAACTGGGCCAGCCAGATCTACCGCGCATCAAGGTGC	4459
QY	1496	GCCAGCTGTGAAGCTGTGCGCGCGCAAGGCCCTTACCGACATCGTGCCTCTGACCG	1555
Db	4460	GCCAGCTGTGAAGCTGTGCTGCGCGCACCAAGGCCCTTACCGAGGTGTGTGCCCTGACCG	4519
QY	1556	AGGAGCCAGCTGAGCTGGCCGAGAACCGCGAGATCCTGGCGAGCCGTGACCGCG	1615
Db	4520	AGGAGCCAGCTGAGCTGGCCGAGAACCGCGAGATCCTGAAAGAGCCCTTGACCGCG	4579

QY	1616	TGTA	CTACGACCCACGACGAGCACTGTGTGGCCGAGATCCAGAAAGCAGGGCCACGACCACT	1675
DB	4580	TGTA	CTACGACCCACGACGAGCACTGTGTGGCCGAGATCCAGAAAGCAGGGCCACGACCACT	4639
QY	1676	GGAC	CTTACCAAGATCTACCAAGAGCCCTTCAAGAAACCTGAAGACCCGCGAAGTACGCCCAAGA	1735
DB	4640	GGAC	CTTACCAAGATCTACCAAGAGCCCTTCAAGAAACCTGAAGACCCGCGAAGTACGCCCGCA	4699
QY	1736	TGCG	CAACGCGCCCAACCAACGACGCTGAAGCAGCTGACCCGAGGCGCTGCAGAAATCGCCA	1795
DB	4700	TGA	AGGGCGGCCACACCAACGACGCTGAAGCAGCTGACCCGAGGCGCTGCAGAAATCGCCA	4759
QY	1796	TGAG	AGCATCGTGTATCTGGGGCAAGACCCCAAGTTCCGCTTCGCCATCCAGAAAGGAGA	1855
DB	4760	CCG	AGCATCGTGTATCTGGGGCAAGACCCCAAGTTCAAGCTGCCATCCAGAAAGGAGA	4819
QY	1856	CCT	GGAGACCTGTGTGACCGACTACTGCGCAGGCCCACTGGATCCCGAGTGGGAGTTTCG	1915
DB	4820	CCT	GGAGGCGCTGTGGACCGACTACTGCGCAGGCCCACTGGATCCCGAGTGGGAGTTTCG	4879
QY	1916	TGA	CACCCCCCCTCTGTGTGAAGCTGTGTACCAAGCTGGAGAGAGGCCCATCATCGCGG	1975
DB	4880	TGA	CACCCCCCCTCTGTGTGAAGCTGTGTGTGTGTACCAAGCTGGAGAGAGGCCCATCATCGCGG	4939
QY	1976	CCG	AGACCTTCTACGTGGACCGCGCCCGCAACCCGCGAGA	2035
DB	4940	CCG	AGACCTTCTACGTGGACCGCGCCCGCAACCCGCGAGA	4999
QY	2036	ACGT	GACCGACCGGGCCGCGCAGAGATCTGTAGCCCTGACCGAGACCAACACGAGAAGA	2095
DB	5000	ACGT	GACCGACCGGGCCGCGCAGAGAGTGTGTCCCTGACCGACCAACACGAGAAGA	5059
QY	2096	CCG	AGCTGACGCCATCCAGCTTGGCCCTTCGAGGACGCGGAGGTGAACATCTGTGA	2155
DB	5060	CCG	AGCTGACGCCATCCAGCTTGGCCCTTCGAGGACGCGGAGGTGAACATCTGTGA	5119
QY	2156	CCG	ACGACGACGATCATCGCCCTTGGGCATCTCAGGCCCGCCGACGAGCGAGCCGAGC	2215
DB	5120	CCG	ACGACGACGATCATCGCCCTTGGGCATCTCAGGCCCGCCGACGAGCGAGCCGAGC	5179
QY	2216	TGGT	GAACGAGATCATCGAGCAGCTGTATCAAGAGAGGAAGGTGTACCTGAGCTGGGTGC	2275
DB	5180	TGGT	GAGCAGATCATCGAGCAGCTGTATCAAGAGAGGAAGGTGTACCTGAGCTGGGTGC	5239
QY	2276	CCG	CCCAAGGGGCATCCGCGCAACGAGCAGATCGACAAGCTGGTGAAGAGGGCATCC	2335
DB	5240	CCG	CCCAAGGGGCATCCGCGCAACGAGCAGATCGACAAGCTGGTGAAGAGGGCATCC	5299
QY	2336	GCA	AGGTGTCTTCTGACCGGCATCGATGGCGGCATCGTGTATCTACCAAGTA	2387
DB	5300	GCA	AGGTGTCTTCTGACCGGCATCGACAAGGGCCCGAGGAGCAGCAGAA	5351

RESULT 7	BD263706	3312 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	BD263706	Improved expression of HIV polypeptides and production of virus-like particles.			
DEFINITION	BD263706	Improved expression of HIV polypeptides and production of virus-like particles.			
ACCESSION	BD263706.1	GI:33073474			
VERSION	JP 2002533124-A/73.	synthetic construct			
KEYWORDS	synthetic construct	artificial sequences.			
SOURCE	artificial sequences.	1 (bases 1 to 3312)			
ORGANISM	1 (bases 1 to 3312)	Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.			
REFERENCE	Improved expression of HIV polypeptides and production of virus-like particles	Patent: JP 2002533124-A 73 08-OCT-2002;			
AUTHORS	CHIRON CORP	OS Artificial Sequence			
TITLE	CHIRON CORP	PN JP 2002533124-A/73			
JOURNAL	PN	JP 2002533124-A/73			
COMMENT	JP 2002533124-A/73				



PD	08-OCT-2002	DB	721	CCAAAGTGGCGCAGCTGGTGGACCTTCCGCGAGCTGAAACAAGCGCACCCAGGACCTTCTGGG	780
PF	30-DEC-1999 JP 2000591193	QY	938	AGGTGCAGCTGGGCATCCCCCAACCCCGCGGCTTGAAGAAGAAGAGCTGACCGTGC	997
PR	31-DEC-1998 US 60/114495, 01-DEC-1999 US 60/168471 PI	DB	781	AGGTGCAGCTGGGCATCCCCCAACCCCGCGGCTTGAAGAAGAAGAGCTGACCGTGC	840
KARIN HARTOG,		QY	998	TGAGCTGGGCGCAGCGCTACTTTCAGCGTCCCTCGACGAGACTTCCGCGAGTACACCG	1057
C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC		DB	841	TGAGCTGGGCGCAGCGCTACTTTCAGCGTCCCTCGACGAGACTTCCGCGAGTACACCG	900
C12N5/10, C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC		QY	1058	CTTTCACATCCCCAGCATCAACAAGAGACCCCGGATCCCGTACCAACAGTGC	1117
PC C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC		DB	901	CTTTCACATCCCCAGCATCAACAAGAGACCCCGGATCCCGTACCAACAGTGC	960
Description of Artificial Sequence: PS(-), protmod, Rtopt(+)		QY	1118	TGCCCGAGGGTGAAGGCGAGCCCGAGCATCTTCCAGAGCAGTACCAAGATCCCTGG	1177
FT source	Location/Qualifiers	DB	961	TGCCCGAGGGTGAAGGCGAGCCCGAGCATCTTCCAGAGCAGTACCAAGATCCCTGG	1020
FT	1. .2312	QY	1178	AGCCCTTCCGCGCCGCAACCCCGAGATCGTGATCTACCA-----GGCCCCCTGTACG	1231
FEATURES	Location/Qualifiers	DB	1021	AGCCCTTCCGCGAGCAGAACCCCGATCGTGATCTACCAAGTACAGTACGACCTGTACG	1080
source	1. .2312	QY	1232	TGGCGAGCGACTCTGAGATCGGCGAGCAGCGCGCCAGATTCGAGAGCTGCCAGACCC	1291
ORIGIN	/organism="synthetic construct"	DB	1081	TGGCGAGCGACTCTGAGATCGGCGAGCAGCGCGCCAGATTCGAGAGCTGCCAGACCC	1140
Query Match	/mol_type="genomic DNA"	QY	1292	TGCTCGCTGGGGCTTCAACACCCCGCAGAAAGCAGCAGAGAGAGCCCGCTTCTGT	1351
Best Local Similarity 93.3%; Pred. No. 3.6e-203;	/db_xref="taxon:32630"	DB	1141	TGCTCGCTGGGGCTTCAACACCCCGCAGAAAGCAGCAGAGAGAGCCCGCTTCTGT	1200
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;		QY	1352	GGATGGGCTACGAGCTGCACCCCGCAGAAAGTGGACCTGCAGGCCATTCGAGCTGCCGAGA	1411
170 GCGGCAAGGAGGCCACACAGATGAGGACTGACAGCGCGGAGTTCACCGAGCAGAACCGGCCA		DB	1201	GGATGGGCTACGAGCTGCACCCCGCAGAAAGTGGACCTGCAGGCCATTCGAGCTGCCGAGA	1260
1 GCGGCGCGAGGACACCAATGAAGATTGCACTGAGAGCAGGCTAATTTCTTCGCG		QY	1412	AGGAGAGCTGGAACCGTGAACGACATCCAGAGCTGGTGGCAGCTGAACTGGCGCCAGCC	1471
230 AGGACCTGGCTTCCCGCAGGCGAAGCGCGGAGTTCACCGAGCAGAGAACCGGCCA		DB	1261	AGGAGAGCTGGAACCGTGAACGACATCCAGAGCTGGTGGCAGCTGAACTGGCGCCAGCC	1320
61 AGGACCTGGCTTCTTCAGGCGAAGCGCGGAGTTTCAGAGCAGCAGACCGCGGCCA		QY	1472	AGATCTACCCCGGATCAAGGTGCGCAGCTGTGCAAGCTGTGCTGCGCGGCAACAAGGCC	1531
290 ACAGCCCAACAGCGCGAGCTCAGGTGCGGCGG-----ACAAACCCCGCAGCAGG		DB	1321	AGATCTACCCCGGATCAAGGTGCGCAGCTGTGCAAGCTGTGCTGCGCGGCAACAAGGCC	1380
121 ACAGCCCAACCGCGGAGCTCAGGTGCGGCGGCGGAGAACACAGCCTGAGCGGAGG		QY	1532	TGACCGAGCATCGTGCCTTACCGAGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGA	1591
344 CCGCGCGGAGCGCAGCGGCAACCTG-----AATTTCCCGCAGATCAACCTGTGGCAGC		DB	1381	TGACCGAGCATCGTGCCTTACCGAGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGA	1440
181 CCGCGCGGAGCGCAGCGGCAACCTGAGCTTCACTTCCCGCAGATCAACCTGTGGCAGC		QY	1592	TGCTCGGAGCGCGCTGCACCGCGCTGTACTACGACCCCGCAGCAGGACCTTGGTGGCCGAGA	1651
398 GCCCCTGTGAGCATCAAGGTGGGCGGCGAGATCAAGAGAGCCCTGTGAGACACCGCGC		DB	1441	TGCTCGGAGCGCGCTGCACCGCGCTGTACTACGACCCCGCAGCAGGACCTTGGTGGCCGAGA	1500
241 GCCCCTGTGAGCATCAAGGTGGGCGGCGAGCTCAAGAGAGCGCTGTGCGACACCGCGC		QY	1652	TGCTCGGAGCGCGCTGCACCGCGCTGTACTACGACCCCGCAGCAGGACCTTGGTGGCCGAGA	1711
458 CCGAGCAGCCGCTCGAGAGATGAGCTGCGCGGAGTGGAGCGCCAGAGATGATCG		DB	1501	TGCTCGGAGCGCGCTGCACCGCGCTGTACTACGACCCCGCAGCAGGACCTTGGTGGCCGAGA	1560
301 CCGAGCAGCCGCTCGAGAGATGAGCTGCGCGGAGTGGAGCGCCAGAGATGATCG		QY	1712	TGCTCGGAGCGCGCTGCACCGCGCTGTACTACGACCCCGCAGCAGGACCTTGGTGGCCGAGA	1771
518 CCGGATCGGGGGCTTCAATCAAGGTGGCGGAGTACGACCCAGATCTCTGATTCGAGATCTCG		DB	1561	TGCTCGGAGCGCGCTGCACCGCGCTGTACTACGACCCCGCAGCAGGACCTTGGTGGCCGAGA	1620
361 CCGGATCGGGGGCTTCAATCAAGGTGGCGGAGTACGACCCAGATCTCTGATTCGAGATCTCG		QY	1772	TGCTCGGAGCGCGCTGCACCGCGCTGTACTACGACCCCGCAGCAGGACCTTGGTGGCCGAGA	1831
578 GCAAGAGGCCATCGGACCGTGTGATCGGCGCGCAGCGCGGAGATCATCTCGCGCGCA		DB	1621	TGCTCGGAGCGCGCTGCACCGCGCTGTACTACGACCCCGCAGCAGGACCTTGGTGGCCGAGA	1680
421 GCAAGAGGCCATCGGACCGTGTGATCGGCGCGCAGCGCGGAGATCATCTCGCGCGCA		QY	1832	TGCTCGGAGCGCGCTGCACCGCGCTGTACTACGACCCCGCAGCAGGACCTTGGTGGCCGAGA	1891
638 ACATGCTGACCCAGTGGGCTGCAACCTTCCCATTCAGCCCGCATCGAGACCGTGC		DB	1681	TGCTCGGAGCGCGCTGCACCGCGCTGTACTACGACCCCGCAGCAGGACCTTGGTGGCCGAGA	1740
481 ACCTGCTGACCCAGATCGGCTGCAACCTTCCCATTCAGCCCGCATCGAGACCGTGC		QY	1892	TGCTCGGAGCGCGCTGCACCGCGCTGTACTACGACCCCGCAGCAGGACCTTGGTGGCCGAGA	1951
698 CCGTGAAGCTGAAGCCCGGATGAAGCGCGCGGAGTGAAGTGAAGTGGCGCGGAGG		DB	1741	TGCTCGGAGCGCGCTGCACCGCGCTGTACTACGACCCCGCAGCAGGACCTTGGTGGCCGAGA	1800
541 CCGTGAAGCTGAAGCCCGGATGAAGCGCGCGGAGTGAAGTGAAGTGGCGCGGAGG		QY	1952	TGCTCGGAGCGCGCTGCACCGCGCTGTACTACGACCCCGCAGCAGGACCTTGGTGGCCGAGA	2011
758 AGAAGATCAAGCCCTGACCCCATCTCGAGGAGATGGAAGAGGAGGCGAGATCACCA		DB			
601 AGAAGATCAAGCCCTGAGATCTGACCGGAGATGGAAGAGGAGGCGAGATCACCA		QY			
818 AGATCGGCGCGAGAACCCCTTACAAACCCCGGCTTGGCGCATCAAGAGAGGAGCAGCA		DB			
661 AGATCGGCGCGAGAACCCCTTACAAACCCCGGCTTGGCGCATCAAGAGAGGAGCAGCA		QY			
878 CCAAGTGGCGCAGCTGGTGGACCTTCCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGGG		DB			



Db	1801	TGGAGAGAGAGCCCATCGTGGGGCCGAGACCTTCTACGTGGAGCGGGCCGCCAACCGCG	1860
Qy	2012	AGACCAAGATCGGCAAGCCGGCTACGTGACCCAGCCGGGGCCGCGAGAGATCGTGAGCC	2071
Db	1861	AGACCAAGCTGGGCAAGCCGGCTACGTGACCCAGCCGGGGCCGCGAGAGATCGTGAGCA	1920
Qy	2072	TGACCGAGAGACCAACCAAGAGCCAGTGCAGGCCATCCAGCTGGCCCTGCAGACA	2131
Db	1921	TCGCCGACACCAACCAAGAGCCAGTGCAGGCCATCCAGCTGGCCCTGCAGACA	1980
Qy	2132	CGGCGAGCGAGGTGAACATCGTGACCCAGACAGCCAGTGCAGGCCATCCAGCGCCC	2191
Db	1981	CGGCGCTGGAGTGAACATCGTGACCCAGACAGCCAGTGCAGGCCATCCAGCGCCC	2040
Qy	2192	AGCCCGACAGAGCGAGCGAGTGTGACCCAGACAGCCAGTGCAGGCCATCCAGAGG	2251
Db	2041	AGCCCGACAGAGCGAGCGAGTGTGACCCAGACAGCCAGTGCAGGCCATCCAGAGG	2100
Qy	2252	AGAGGTGTACTGAGTGGGTGCCGCCGCCACAGAGGCGATCGGCGGCAAGAGCAGATCG	2311
Db	2101	AGAGGTGTACTGAGTGGGTGCCGCCGCCACAGAGGCGATCGGCGGCAAGAGCAGTGG	2160
Qy	2312	ACAGCTGGTGAGCAAGGGGATCCGCAAGGTGTGTTCTTGGAGCGGATCGATGGGGCA	2371
Db	2161	ACAGCTGGTGAGCGCGGCGATCCGCAAGGTGTGTTCTTGGAGCGGATCGATGGGGCA	2220
Qy	2372	TCGTGATCTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC	2431
Db	2221	TCGTGATCTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC	2280
Qy	2432	AAAAGCTTCCGGGGCTAGCACCGGTGAATTC	2463
Db	2281	AAAAGCTTCCGGGGCTAGCACCGGTGAATTC	2312
RESULT 8			
AR373389			
LOCUS			
DEFINITION			
Sequence 84 from patent US 6602705.			
ACCESSION			
AR373389			
VERSION			
AR373389.1			
KEYWORDS			
GI:40075492			
SOURCE			
Unknown.			
ORGANISM			
Unknown.			
REFERENCE			
1 (bases 1 to 2312)			
Barnett, S.W., Megede, J., Greer, C. and Selby, M.			
Expression of HIV polypeptides and production of virus-like			
particles			
Patent: US 6602705-A 84 05-AUG-2003;			
JOURNAL			
Location/Qualifiers			
1..2312			
/organism="unknown"			
/mol_type="genomic DNA"			
ORIGIN			
Query Match 82.2%; Score 2025.2; DB 6; Length 2312;			
Best Local Similarity 93.3%; Pred. No. 3.6e-203;			
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;			
Qy	170	GGGCAAGAGGAGGCGCCAGATGAGGAGTGCACCGAGCGGCGAGGCACTTCTTCGCG	229
Db	1	GGGCGGGGAGGACACCAAAATGAAGATTGCACTGAGAGACAGGCTAATTTCTTCGCG	60
Qy	230	AGGAGCTGGCCCTTCCCGGAGGCGGCGGAGTTCGCCAGCGAGCAGAACCGCGCA	289
Db	61	AGGAGCTGGCCCTTCTGCGAGGCGGAGGCGGCGGAGTTCAGCAGCAGGACCGCGCA	120
Qy	290	ACAGCCCGACAGCGCGAGCTGCGAGTTCGCGGCGG-----ACAAACCGCGAGG	343
Db	121	ACAGCCCGACCGCGCGAGCTGCGAGTTCGCGGCGGAGGAAACACAGCTGAGCGAGG	180
Qy	344	CCGGCGCGAGCGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	397

Db	181	CGGGCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	240
Qy	398	GGCCCTGGTGGATCAAGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	457
Db	241	GGCCCTGGTGGATCAAGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	300
Qy	458	CGGAGCACCGCTGGTGGAGGAGATGAGCTGCCCGGCAAGTGAAGCAAGAGATGATCG	517
Db	301	CGGAGCACCGCTGGTGGAGGAGATGAGCTGCCCGGCAAGTGAAGCAAGAGATGATCG	360
Qy	518	GGGGCATCGGGGGTTCATCAAGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	577
Db	361	GGGGCATCGGGGGTTCATCAAGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	420
Qy	578	GCAGAGAGCGCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG	637
Db	421	GCAGAGAGCGCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG	480
Qy	638	ACATCTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	697
Db	481	ACCTGCTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	540
Qy	698	CGGTGAAGTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG	757
Db	541	CGGTGAAGTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG	600
Qy	758	AGAAGATCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	817
Db	601	AGAAGATCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	660
Qy	818	AGATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	877
Db	661	AGATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	720
Qy	878	CCAAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	937
Db	721	CCAAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	780
Qy	938	AGGTGAGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	997
Db	781	AGGTGAGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	840
Qy	998	TGAGCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1057
Db	841	TGAGCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	900
Qy	1058	CCTTACCATCCCGAGCATCAACAGGAGACCCCGGCGATCCGCTACAGTACAACTGCG	1117
Db	901	CCTTACCATCCCGAGCATCAACAGGAGACCCCGGCGATCCGCTACAGTACAACTGCG	960
Qy	1118	TGCCCCAGGGCTGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1177
Db	961	TGCCCCAGGGCTGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1020
Qy	1178	AGCCCTTCCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1231
Db	1021	AGCCCTTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1080
Qy	1232	TGCGCAGCGCATCTGGAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1291
Db	1081	TGCGCAGCGCATCTGGAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1140
Qy	1292	TGCTGGCTGGGGCTTCCACACCCCGGCAAGAGCAGCAGAGGAGGAGGAGGAGGAGGAGG	1351
Db	1141	TGCTGGCTGGGGCTTCCACACCCCGGCAAGAGCAGCAGAGGAGGAGGAGGAGGAGGAGG	1200
Qy	1352	GGATGGCTACGAGCTGCGACCCCGGCAAGTGGACCGTGCAGGCCATTCGAGCTGCCGAGA	1411
Db	1201	GGATGGCTACGAGCTGCGACCCCGGCAAGTGGACCGTGCAGGCCATTCGAGCTGCCGAGA	1260
Qy	1412	AGGAGCTGCGAGCTGCGACCGCATCCAGAGCTGCTGGGAGCTGAGCTGCGGCGGCGGCGG	1471

1261 AGGACAGCTGGACCGTGAACGACATCCAGAAAGCTGGTGGCAAGCTGAATCGGGCCAGCC 1320  
1472 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGCGCGCCAGGCGCC 1531  
1321 AGATCTACCGCGGCATCAAGGTGAAGCAGTGTGCAAGCTGTGCGCGCGCCAGGCGCC 1380  
1532 TGACCGACATCTGTCGCCCTGACCGAGAGCGCCAGCTGAGCTGGCGCGGAGAACCGCGAGA 1591  
1381 TGACCGAGGTGATCCCGCTGACCGAGAGCGCCAGCTGGAGCTGGCGCGGAGAACCGCGAGA 1440  
1592 TCCTGCGGAGCGCGTGCACCGGTGTACTACGACCCAGCAGGACCTGGTGGCGCGAGA 1651  
1441 TCCTGAAGAGCGCGTGCACGAGGTGTACTACGACCCAGCAGGACCTGGTGGCGCGAGA 1500  
1652 TCAGAGAGCGCGCCAGCAGAGTGGACCTACAGATCTACAGAGAGCCCTTCAAGAAC 1711  
1501 TCAGAGAGCGCGCCAGGCGCAGTGGACCTACAGATCTACAGAGAGCCCTTCAAGAAC 1560  
1712 TGAAGACCGGCAAGTACGCAAGATGCGCAGCGCCACACCAACGACGCTGAAGCAGCTGA 1771  
1561 TGAAGACCGGCAAGTACGCGCGCATGCGCGCGCCACACCAACGACGCTGAAGCAGCTGA 1620  
1772 CCGAGGCGCGTGCAGAAAGTCCGATGAGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGT 1831  
1621 CCGAGGCGCGTGCAGAAAGTCCGATGAGAGAGCATCTGTGATCTGGGGCAAGATCCCAAGT 1680  
1832 TCCGCTGCGCATCCAGAGAGAGACCTGGAGACCTGTGTGACCGCTACTGTGACGCGCA 1891  
1681 TCAAGCTGCGCATCCAGAGAGAGACCTGGAGGCGCTGTGTGATGAGTACTGTGACGCGCA 1740  
1892 CTTGATCCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGTTGTAAGCTGTGTACGAGC 1951  
1741 CTTGATCCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGTTGTAAGCTGTGTACGAGC 1800  
1952 TGGAGAGGAGCCCATCATCGCGCGCAGACCTTACGTGGAGCGCGCGCCCAACCGCG 2011  
1801 TGGAGAGGAGCCCATCATCGTGGCGCGCAGACCTTACGTGGAGCGCGCGCCCAACCGCG 1860  
2012 AGACCAAGTCCGCAAGCGCGCTACGTGACCGACCGCGCGCGCGCAGAGATCGTAGCC 2071  
1861 AGACCAAGTCCGCAAGCGCGCTACGTGACCGACCGCGCGCGCGCAGAGTGTGTAGCA 1920  
2072 TGACCGAGACCAACCAAGAGACCGAGCTGTGAGACCGATCCAGCTGGCGCTGCGAGCA 2131  
1921 TCGCGGACACCAACCAAGAGACCGAGCTGTGAGACCGATCCAGCTGGCGCTGCGAGCA 1980  
2132 CCGGCGAGCGAGTGAACATCTGTGACCGACAGCCAGTACGCGCTGGGCAATCATCCAGGCC 2191  
1981 CCGGCTCGAGGTGAACATCTGTGACCGACAGCCAGTACGCGCTGGGCAATCATCCAGGCC 2040  
2192 AGCCCGACAGAGCGAGCGAGCTGTGAGACCGATCATCGACAGCTGATCAAGAGG 2251  
2041 AGCCCGACAGAGCGAGCGAGCTGTGAGACCGATCATCGACAGCTGATCAAGAGG 2100  
2252 AGAAGGTGTACTGAGCTGGGTGCGCGCCCAAGAGGATCGCGCGCAACGAGCAGATCG 2311  
2101 AGAAGGTGTACTGAGCTGGGTGCGCGCCCAAGAGGATCGCGCGCAACGAGCAGTGG 2160  
2312 ACAAGCTGGTGAAGGCGATCCGCAAGGTGTCTTCTGGAGCGGATCGATGGCGGCA 2371  
2161 ACAAGCTGGTGAAGGCGATCCGCAAGGTGTCTTCTGGAGCGGATCGATGGCGGCA 2220  
2372 TCGTGATCTACAGTACATGACGACCTGTACGTGGGCGAGCGCGCGCTAGGATCGATT 2431  
2221 TCGTGATCTACAGTACATGACGACCTGTACGTGGGCGAGCGCGCGCTAGGATCGATT 2280  
2432 AAAAGCTTCCGGGCTAGCACCGGTGAATTC 2463  
2281 AAAAGCTTCCGGGCTAGCACCGGTGAATTC 2312

RESULT 9  
BD263705

LOCUS BD263705 2300 bp DNA linear PAT 17-JUL-2003  
DEFINITION Improved expression of HIV polypeptides and production of  
virus-like particles.  
ACCESSION BD263705  
VERSION BD263705.1 GI:33073473  
KEYWORDS JP 2002533124-A/72.  
SOURCE synthetic construct  
ORGANISM synthetic construct  
artificial sequences.  
REFERENCE 1 (bases 1 to 2300)  
AUTHORS Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,  
Greer,C., Selby,M. and Walker,C.  
TITLE Improved expression of HIV polypeptides and production of  
virus-like particles  
JOURNAL Patent: JP 2002533124-A 72 08-OCT-2002;  
CHIRON CORP  
COMMENT OS Artificial Sequence  
PN JP 2002533124-A/72  
PD 08-OCT-2002  
PF 30-DEC-1999 JP 2000591193  
PR 31-DEC-1998 US 60/114495, 01-DEC-1999 US 60/168471 PI  
SUSAN BARNETT, JAN ZUR MEGEDE, INDRESH SRIVASTAVA, YING LIAN, PI  
KARIN HARTOG  
PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER PC  
C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC  
C12N5/10,  
PC C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61X3/02 CC  
Description of Artificial Sequence: FS(-). protomod.Rtopt.YMMH PH  
Key Location/Qualifiers  
FT source 1..2300  
FT /organism='Artificial Sequence'.  
FEATURES  
source Location/Qualifiers  
1..2300  
/organism='synthetic construct'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32630'

ORIGIN  
Query Match 82.0%; Score 2019.2; DB 6; Length 2300;  
Best Local Similarity 93.2%; Pred. No. 1.5e-202;  
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;  
QY 170 GCGGCAAGGAGGGCCACAGATGAAGGACTGACCGAGCGCGCAGGCCAACCTTTTCGCG 229  
DB 1 GCGGCGCGGAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAATTTCTTCGCG 60  
QY 230 AGGACTGCGCTTCCCGAGGGCAAGGCCCGCGAGTTCCCGAGCGAGCAACCGCGCA 289  
DB 61 AGGACTGCGCTTCCCGAGGGCAAGGCCCGCGAGTTCCAGCGAGCGAGACCGCGCA 120  
QY 290 ACAGCCCGACCGCGCGAGCTGCGAGTGCAGGTGCGCGGCGCGAGCAACAGCGCGAGG 343  
DB 121 ACAGCCCGACCGCGCGAGCTGCGAGTGTGGGCGCGCGAGCAACAGCGCGAGG 180  
QY 344 CCGGCGCGGAGCGCGCGAGCGCACCTG-----ACCTTCCCGAGATCACTTGTGGCAGC 397  
DB 181 CCGGCGCGCGCGCGCGAGCGCACCTGAGCTTCACTTCCCGAGATCACTTGTGGCAGC 240  
QY 398 GCGGCTGTGAGCATCAAGTGGCGCGCGAGTCAAGGAGCGCGCTGTGACACCGCGCG 457  
DB 241 GCGGCTGTGAGCATCAAGTGGCGCGCGAGTCAAGGAGCGCGCTGTGACACCGCGCG 300  
QY 458 CCGAGCACACCGTGTGAGGAGATGAGCTTCCCGCGCAAGTGGAGCGCGCAAGATGATCG 517  
DB 301 CCGAGCACACCGTGTGAGGAGATGAACTTCCCGCGCAAGTGGAGCGCGCAAGATGATCG 360  
QY 518 GCGGCTGTGAGCATCAAGTGGCGCGCGAGTGCAGTACAGCATCTGTGAGATCTGCG 577  
DB 361 GCGGCTGTGAGCATCAAGTGGCGCGCGAGTGCAGTACAGCATCTGTGAGATCTGCG 420  
QY 578 GCAAGAGGCGCATCGGCGCGCGCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 637  
DB 421 GCCACAGGCGCATCGGCGCGCGCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 480

Qy	1718	CGCGCAAGTACGCCAAGATGCGCACCGGCCACACACGAGCTGAGACAGCTGACCCGAGG	1777
Db	1555	CGGCAAGTACGCCCGCATGCGCGCGGCCACACCAACGAGCTGAGACAGCTGACCCGAGG	1614
Qy	1778	CGGTGCAGAGAGATCGCCATGTGAGAGAGATCTGGGCGAAGACCCCAAGTTCCGCC	1837
Db	1615	CGGTGCAGAGAGGTGAGCACCGAGAGCATCTGTGATCTGGGCGAAGATCCCCAAGTTCAGC	1674
Qy	1838	TGCCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTPACTGGCAGGGCCACTGGA	1897
Db	1675	TGCCCCATCCAGAAGGAGACCTGGGAGGCTGTGGATGCGACTCTGGCAGGCGCACTGGA	1734
Qy	1898	TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA	1957
Db	1735	TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA	1794
Qy	1958	AGGAGCCCATCATCGCGGCGGAGACCTTCTAGCTGGAACGGCGCCCGCCAAACCCGAGACCA	2017
Db	1795	AGAGGCCCATCTGTGGGCGCGGAGACCTTCTTAGTGAACGGCGCCCGCCAAACCCGAGACCA	1854
Qy	2018	AGATCGGCAAGCCCGCTACGTGAACGACCGGGGCGCGGAGAGATCTGTGAGCTGACCG	2077
Db	1855	AGCTGGGCAAGCCCGCTACGTGAACGACCGGGGCGCGGAGAGGTGTGTGACATCGCCG	1914
Qy	2078	AGACCACCAACAGAGAACCGAGCTCAGCGGCATCCAGCTGGCCCTGCAGGACAGCGGCA	2137
Db	1915	ACACCAACCAACAGAGAACCGAGCTCAGCGGCATCCACCTGGCCCTGCAGGACAGCGGCC	1974
Qy	2138	CGGAGGTGAACATCGTGAACCGACAGCGAGTACGCCCTGGGCACTCATCCAGGCCACGCCG	2197
Db	1975	TGAGGTGAACATCGTGAACCGACAGCAAGTACGCCCTGGGCACTCATCCAGGCCACGCCG	2034
Qy	2198	ACAAGACGAGACGAGCTTGGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGAGG	2257
Db	2035	ACAAGACGAGACGAGCTTGGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGAGG	2094
Qy	2258	TGTACTGAGCTGGGTGCCGCCCCACAAAGGGCATCGGCGCAACGAGCAGATCGACAAGC	2317
Db	2095	TGTACTTGGGCTGGGTGCCGCCCCCAAGGGCATCGGCGCAACGAGCAGCTGGAACAGC	2154
Qy	2318	TGTTGACAAAGGGCATCCGCAAGGTGTGTCTCTTGGACGGCATCTGATGGCGGCATCGTGA	2377
Db	2155	TGTTGAGCGCGGCATCCGCAAGGTGTGTCTCTTGGACGGCATCGATGGCGGCATCGTGA	2214
Qy	2378	TCTTACCAAGTACATGAGCAGACTGTATCGTGGGACAGCGGCGGCCCTTAGGATCGATTAAAGC	2437
Db	2215	TCTTACCAAGTACATGAGCAGACTGTATCGTGGGACAGCGGCGGCCCTTAGGATCGATTAAAGC	2274
Qy	2438	TTCCCCGGGGCTAGCACCGGTGAATTC	2463
Db	2275	TTCCCCGGGGCTAGCACCGGTGAATTC	2300

RESULT 10  
 AR373388  
 LOCUS  
 DEFINITION  
 DEFINITION Sequence 83 from patent US 6602705.  
 AR373388  
 ACCESSION  
 AR373388.1 GI:40075491  
 VERSION  
 KEYWORDS  
 SOURCE  
 SOURCE Unknown.  
 ORGANISM  
 ORGANISM Unclassified.  
 1 (bases 1 to 2300)  
 REFERENCE  
 BARNETT, S.W., McEDE, J., GREER, C. and SELBY, M.  
 AUTHORS  
 Expression of HIV polypeptides and production of virus-like  
 TITLE  
 Patent: US 6602705-A 83 05-AUG-2003;  
 JOURNAL  
 Location/Qualifiers  
 FEATURES  
 1..2300  
 /organism="unknown"  
 /mol\_type="genomic DNA"  
 source  
 ORIGIN

Query Match 82.0%; Score 2019.2; DB 6; Length 2300;			
Best Local Similarity 93.2%; Pred. No. 1.5e-202;			
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;			
QY	170	GGGCAAGAGGCGCCACCAAGATGAAGACTGACCGAGCGCCAGCCAACTTCTTCGCG	229
DB	1	GGGCGCGGAGGACCAACAAATGAAGATTGATGAGAGACAGCTAATTTCTTCGCG	60
QY	230	AGGACTTGGCTTCCCGCAGGCGAAGCCCGCAGTTTCCCGAGCAGAGAACCGCGCCA	289
DB	61	AGGACTTGGCTTCTGAGGCGCAAGCCCGCAGTTTCCCGAGCAGAGAACCGCGCCA	120
QY	290	ACAGCCCAACAGCGCGAGCTGACGTGCGCGG-----ACAACCCCGCAGCGAG	343
DB	121	ACAGCCCAACCGCGCGAGCTGACGTGCGCGGCGGAGAACACAGCTTGAAGCG	180
QY	344	CGGCGCGAGGCGCCAGGCGCACTTG-----AATTTCCCGCAGATCACCTTGGCAGC	397
DB	181	CGGCGCGAGCGCGCAGGCGCACTTGAGCTTCAACTTCCCGCAGATCACCTTGGCAGC	240
QY	398	GGCCCTTGGTGAAGCATCAAGGTGGCGCGCAGATCAAGGAGGCGCTGCTGACACCGCG	457
DB	241	GGCCCTTGGTGAACATCAAGTGGCGCGCAGATCAAGGAGGCGCTGCTGACACCGCG	300
QY	458	CGGAGACACCGTGTGAGGAGATGAGCTTGCCTGCGCAAGTGAAGCCCAAGATGATCG	517
DB	301	CGGAGACACCGTGTGAGGAGATGAACTTGCCTGCGCAAGTGAAGCCCAAGATGATCG	360
QY	518	GGGCGATCGGCGCTTCAATCAAGGTGGCGCGCAGATCAAGGAGGCGCTGCTGACATCTCG	577
DB	361	GGGCGATCGGCGCTTCAATCAAGGTGGCGCGCAGATCAAGGAGGCGCTGCTGACATCTCG	420
QY	578	GGCAAGGCGCATCGGCACTGTGTGATCGGCGCCCAACCGCGTGAACATCATCGGCGCA	637
DB	421	GGCAAGGCGCATCGGCACTGTGTGATCGGCGCCCAACCGCGTGAACATCATCGGCGCA	480
QY	638	ACATCTGACACCGAGCTGGCGCTGACCTTGAACCTTCCCGCATCAGCGCCCATCGAGACCGTGC	697
DB	481	ACCTCTGACACCGAGATCGGCTGACCTTGAACCTTCCCGCATCAGCGCCCATCGAGACCGTGC	540
QY	698	CGGTGAAGCTGAAGCGGCGGATGACCGGCCCAAGGTCAAGCTGCGCCCTGACCGGAG	757
DB	541	CGGTGAAGCTGAAGCGGCGGATGACCGGCCCAAGGTCAAGCTGCGCCCTGACCGGAG	600
QY	758	AGAAGATCAAGGCGCTGACCGCATCTGCGAGGAGATGGAAGAGGCGCAAGATCACCA	817
DB	601	AGAAGATCAAGGCGCTGCTGGAGATCTGCACCGAGATGGAAGAGGCGCAAGATCACCA	660
QY	818	AGATCGGCGCGAGAACCCCTACACACCCCGCTGTTGCGCATCAAGAAAGGACGCA	877
DB	661	AGATCGGCGCGAGAACCCCTACACACCCCGCTGTTGCGCATCAAGAAAGGACGCA	720
QY	878	CCAAGTGGCGCAAGCTGTGAGCTTCCGCGAGCTGAAACAGCGCACCCAGGACTTCTGGG	937
DB	721	CCAAGTGGCGCAAGCTGTGAGCTTCCGCGAGCTGAAACAGCGCACCCAGGACTTCTGGG	780
QY	938	AGGTGACGTGGCATCCCCACCCCGCGCTGAGAGAGAGAGAGGTGACCGTGC	997
DB	781	AGGTGACGTGGCATCCCCACCCCGCGCTGAGAGAGAGAGAGGTGACCGTGC	840
QY	998	TGGAGTGGCGCAAGCTTACTGAGCTGCGCCCTGGAAGAGGACTTCCGCAAGTACACCG	1057
DB	841	TGGAGTGGCGCAAGCTTACTGAGCTGCGCCCTGGAAGAGGACTTCCGCAAGTACACCG	900
QY	1058	CTTTCACCATCCCGAGCTACACAGCAGAGCCCGCGCATCCGCTACCACTACAACTGC	1117
DB	901	CTTTCACCATCCCGAGCTACACAGCAGAGCCCGCGCATCCGCTACCACTACAACTGC	960
QY	1118	TGCCCCAGGCTGGAAGGCGAGCCCGCAGCATCTTCCAGAGCAGCATGACCAAGATCTTGG	1177
DB	961	TGCCCCAGGCTGGAAGGCGAGCCCGCAGCATCTTCCAGAGCAGCATGACCAAGATCTTGG	1020
QY	1178	AGCCCTTCCGCGCCCGCAACCCCGAGATGCTGATCTACAGGCGCCCGCTGTACGTGGCA	1237
DB	1021	AGCCCTTCCGCAAGAGAACCCCGACATCGTATCTTACAGGCGCCCGCTGTACGTGGCA	1080
QY	1238	CGGACTTGGAGATCGGCGAGCACCGCGCAAGATCGAGGAGCTGCGAGCACCTGCTGC	1297
DB	1081	CGGACTTGGAGATCGGCGAGCACCGCGCAAGATCGAGGAGCTGCGAGCACCTGCTGC	1140
QY	1298	GCTGGGCTTTCACCAACCCCGCAGCAAGAACACCAAGAGAGCCCGCTTCTGTGGATGG	1357
DB	1141	GCTGGGCTTTCACCAACCCCGCAGCAAGAACACCAAGAGAGCCCGCTTCTGTGGATGG	1198
QY	1358	GCTACGAGCTGCAACCCCGCAGCAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAGGAGA	1417
DB	1199	-----CGAGCTTGCACCCCGCAGCAAGTGGACCGTGCAGCCCATCGTCCCGAGAGGAGA	1254
QY	1418	GCTGGAACGCTGAACGACATCCAGAACTGGTGGGCAAGCTGAACCTGGGCGAGCCAGATCT	1477
DB	1255	GCTGGAACGCTGAACGACATCCAGAACTGGTGGGCAAGCTGAACCTGGGCGAGCCAGATCT	1314
QY	1478	ACCCGCGCATCAAGGTGCGCGCAGCTGTGCAAGCTGTGCGCGCGCGCCCAAGGCCCTGACCG	1537
DB	1315	ACGCGCGCATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGCGCACCAAGGCCCTGACCG	1374
QY	1538	ACATCGTGGCCCTGACCGAGGAGCGCAGCTGAGAGCTGCGCGAGAACCGCGAGATCTGCG	1597
DB	1375	AGGTGATCTCCCTGACCGAGGAGCGCAGCTGAGAGCTGCGCGCGAGAACCGCGAGATCTGCG	1434
QY	1598	CGGAGCCGCTGACCGCGCTGTACTAGCACCCAGCAAGACCTGTGGCGCGAGATCTCAGA	1657
DB	1435	AGGAGCCGCTGACCGAGGCTGTACTAGCACCCAGCAAGACCTGTGGCGCGAGATCTCAGA	1494
QY	1658	AGGAGGCGCAGCAGCTGAGACCTACAGATCTTACAGAGGCGCTTCAAGAACTCTGAAGA	1717
DB	1495	AGGAGGCGCAGGCGCAGCTGAGACCTACAGATCTTACAGAGGCGCTTCAAGAACTCTGAAGA	1554
QY	1718	CCGCAAGTACGCGCAGATCGCACCCCGCACACCAACGACCTGGAAGCAGCTGACCGAGG	1777
DB	1555	CCGCAAGTACGCGCAGATCGCACCCCGCACACCAACGACCTGGAAGCAGCTGACCGAGG	1614
QY	1778	CGGTGAGAGATCGCATGAGAGCATGTGATCTGGGGCAAGACCCCGCAAGTTCGCGC	1837
DB	1615	CGGTGAGAGATCGCATGAGAGCATGTGATCTGGGGCAAGACCCCGCAAGTTCGCGC	1674
QY	1838	TGCCATCTCCAGAAAGGAGACCTTGGGAGACCTGTGTGAGACCTACTTGGCAGGCGCAGCTGA	1897
DB	1675	TGCCATCTCCAGAAAGGAGACCTTGGGAGACCTGTGTGAGTGTGAGTGTGAGAGGCGCAGCTGA	1734
QY	1898	TCCCGGAGTGGAGTTCGTGAACACCCCGCCCTGTGTAAGCTGTGTTACGAGCTGGA	1957
DB	1735	TCCCGGAGTGGAGTTCGTGAACACCCCGCCCTGTGTAAGCTGTGTTACGAGCTGGA	1794
QY	1958	AGGAGCCCATCATCGCGCGCAGACCTTCTAGTGGACGCGCGCGCCCAACCCCGAGAGCA	2017
DB	1795	AGGAGCCCATCATCGCGCGCGCAGACCTTCTAGTGGACGCGCGCGCCCAACCCCGAGAGCA	1854
QY	2018	AGATCGCAAGGCGGCTAGTGACCGAGCGGGCGCGGAGAGATCTGTGAGCTGAGCG	2077
DB	1855	AGTGGGCAAGGCGGCTACGTGACCGAGCGGGCGCGGAGAGGTTGTGAGCATCGCGC	1914
QY	2078	AGACCAACCAACAGAGACCGAGCTCAGGCCATCCAGCTGGCCCTTGCAGGACAGCGCA	2137
DB	1915	ACACCAACCAACAGAGACCGAGCTCAGGCCATCCAGCTGGCCCTTGCAGGACAGCGGC	1974
QY	2138	CGGAGTGAACATCTGTGACCGAGCAGCGAGTACGCTTGGGCGCATCATCAGGCGCGCGC	2197
DB	1975	TGGAGTGAACATCTGTGACCGAGCAGCGAGTACGCTTGGGCGCATCATCAGGCGCGCGC	2034
QY	2198	ACAAGAGCAGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTGATCAAGAAAGGAGG	2257
DB	2035	ACAAGAGCAGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTGATCAAGAAAGGAGG	2094
QY	2258	TGTACTGAGCTGGTGGCGCGCCCAACAGGGCATCGGCGCGCAACGAGCAGATCTGACAGC	2317

Db 2095 TGTACTGGCTGGTGGCCGACCAAGGCGATCGGCGCAACGAGCGGTGACACAGC 2154  
QY 2318 TGTGTAGCAAGGCGATCCGCAAGGTGCTGTTCTGTGACGCGCATCGATCGCGGCGATCGTGA 2377  
Db 2155 TGTGTAGCGCGCGCATCCGCAAGGTGCTGTTCTGTGACGCGCATCGATCGCGGCGATCGTGA 2214  
QY 2378 TCTACCAAGTACATGAGCAAGGCGTGTAGTGTGGGAGCGCGGCGGCGCTAGGATCATTAAGAAGC 2437  
Db 2215 TCTACCAAGTACATGAGCAAGGCGTGTAGTGTGGGAGCGCGGCGGCGCTAGGATCATTAAGAAGC 2274  
QY 2438 TTCCCGGGGCTAGCACGGGTGAATTC 2463  
Db 2275 TTCCCGGGGCTAGCACGGGTGAATTC 2300

RESULT 11  
AX427936  
LOCUS AX427936 9788 bp DNA linear PAT 20-JUN-2002  
DEFINITION Sequence 174 from Patent WO0232943.  
ACCESSION AX427936  
VERSION AX427936.1 GI:21538023  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
artificial sequences.

REFERENCE  
1. Huang, Y. and Nabel, G.J.  
Modifications of hiv env, gag, and pol enhance immunogenicity for  
genetic immunization  
Patent: WO 0232943-A 174 25-APR-2002;  
GOVERNMENT OF THE UNITED STATES (US)  
Location/Qualifiers  
1. 9788  
/organism="synthetic construct"  
/mol\_type="unassigned dna"  
/db\_xref="taxon:32630"  
/note="plasmid pVR1012x/s containing HIV genes"

ORIGIN  
Query Match 82.0%; Score 2019.2; DB 6; Length 9788;  
Best Local Similarity 91.6%; Pred. No. 1e-202;  
Matches 2190; Conservative 0; Mismatches 178; Indels 24; Gaps 4;  
QY 14 TGCCGAGGCGCATGAGCAGGCGCACAGGCGCAACATCTGTATGAGCGCGAGCAACTGCA 73  
Db 2961 TGCCGAGGCGCATGAGCAGGCGTGAACACACCAACATCATGATGAGCGGCGCAACTGCA 3020  
QY 74 AGGCGCCCAAGCGCATCATCAAGTGTCTCACTGCGGCAAGGCGGCACATCGCCGCA 133  
Db 3021 AGGCGCCCAAGCGCATCATCAAGTGTCTCACTGCGGCAAGGCGGCACATCGCCGCGCA 3080  
QY 134 ACTGCGCGCGCCCGCCCAAGAGGGGTGCTGGAAGTGGCGCAAGAGGGCGCCACAGATGA 193  
Db 3081 ACTGCGCGCGCCCGCCCAAGAGGGGTGCTGGAAGTGGCGCAAGAGGGCGCCACAGATGA 3140  
QY 194 AGGACTGCAAGCGCGCGCGCGCGCAACTTCTTCCGCGGAGCGTGGCTTCCCGCGGCGCA 253  
Db 3141 AGGACTGCAAGCGCGCGCGCGCGCGCAACTTCTTCCGCGGAGCTGGCGCTTCCCGCGGCGCA 3194  
QY 254 AGGCGCGCGAGTTCCCG 313  
Db 3195 AGGCGCGCGAGTTTCTTCCAGAGCGAGCCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3254  
QY 314 AGGTGCGCGG-----CGACACCG 364  
Db 3255 AGGTGCGCGGAGAGACACACTCTCTCTGAGAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 3314  
QY 365 ---CCCTGAACTTCCCG 421  
Db 3315 TATCCCTTATGCTTCCCTCAGATCACTCTTTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3374  
QY 422 GCGCGCGAGATCAAGGAGGCG 481

Db 3375 GGGGCGAGCTGAGGAGGCGCTTTAGACACCGCGCGCGAGCACCGTCTGGAGGAGA 3434  
QY 482 TGAGCTGCGCGCGCAAGTGGAAAGCCCAAGATGATGCGCGCGAGATCGCGGCTTTCATCAAGG 541  
Db 3435 TGAACCTGCGCGCGCGCTGGAAAGCCCAAGATGATGCGCGCGAGTCTGGCGGCTTTCATCAAGG 3494  
QY 542 TCGCGCGAGTACGACAGATCTGTATCGAGATCTCGCGCAAGAGGCGCATCGGCGACCGTGC 601  
Db 3495 TGGGCGAGTACGACAGATCTGTATCGAGATCTCGCGCGCAAGAGGCGCATCGGCGACCGTGC 3354  
QY 602 TGATCG 661  
Db 3555 TGGTGGCG 3614  
QY 662 CCCTGAATCTTCCCG 721  
Db 3615 CCCTGAATCTTCCCG 3674  
QY 722 AGGCG 781  
Db 3675 ACGGCG 3734  
QY 782 TCTGCGAGGATGAGAGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 841  
Db 3735 TCTGCGAGGATGAGAGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3794  
QY 842 ACACCG 901  
Db 3795 ACACCG 3854  
QY 902 TCGCGAGCTGAACAAAGCG 961  
Db 3855 TCGCGAGCTGAACAAAGCG 3914  
QY 962 CCG 1021  
Db 3915 CCG 3974  
QY 1022 GCGTGCCTCGGAGGAGCTTCGCGCAAGTACACCGCGCTTCCACATCCCGCGCGCGCGCGCGCG 1081  
Db 3975 GCGTGCCTCGGAGGAGCTTCGCGCAAGTACACCGCGCTTCCACATCCCGCGCGCGCGCGCGCG 4034  
QY 1082 ACAGACCG 1141  
Db 4035 ACAGACCG 4094  
QY 1142 CCAGCATCTTCCAGAGCGCATGACCAAGATCTCTGAGCGCGCTTCCCGCGCGCGCGCGCGCGCG 1201  
Db 4095 CCGCATCTTCCAGTGCAGCATGACCAAGATCTCTGAGCGCGCTTCCCGCGCGCGCGCGCGCGCG 4154  
QY 1202 AGATCGTATCTACCA-----GGCG 1255  
Db 4155 ACATCGTATCTACCAAGTATGAGCACCTGTACGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 4214  
QY 1256 AGCAGCG 1315  
Db 4215 AGCAGCG 4274  
QY 1316 CCGACAAGAGCACCAAGAGGAGCG 1375  
Db 4275 CCGACAAGAGCACCAAGAGGAGCG 4334  
QY 1376 ACAAGTGCAGCG 1435  
Db 4335 ACAAGTGCAGCG 4394  
QY 1436 TCCAGAGCTGTGGGCG 1495  
Db 4395 TCCAGAGCTGTGGGCG 4454  
QY 1496 GCGAGCTGTGAGCG 1555

Db 4455 GCAGCTGTGCAAGTGTCTGCGCGCACCAAGAGCCCTGACCGAGTGTGTGCTGACCG 4514  
 QY 1556 AGAGGCCGAGCTGGAGTGTGCGCGAGAACCGGAGATCTCTCGGAGCCGCTGTGACGCGG 1615  
 Db 4515 AGAGGCCGAGCTGGAGTGTGCGCGAGAACCGGAGATCTCTGAGAGCCGCTGTGACGCGG 4574  
 QY 1616 TGCTACTAGACCGCAGCAGGACCTGTGCGCGAGATCTCAGAGCAGGCGCACGACGAGT 1675  
 Db 4575 TGCTACTAGACCGCAGCAGGACCTGTGCGCGAGATCTCAGAGCAGGCGCACGAGGCGGCGAGT 4634  
 QY 1676 GGACCTTACAGATCTTACAGAGGCGCTTCAAGAACCTTGAAGACCGGCAAGTGTGCGCAAGA 1735  
 Db 4635 GGACCTTACAGATCTTACAGAGGCGCTTCAAGAACCTTGAAGACCGGCAAGTGTGCGCA 4694  
 QY 1736 TCGCGACCGCCACACACAGCAGTGAAGAGCTGACCGAGGCGCTGACAGAGATGCGCA 1795  
 Db 4695 TGAAGGCGCCACACACAGCAGTGAAGAGCTGACCGAGGCGCTGACAGAGATGCGCA 4754  
 QY 1796 TGGAGAGCATCTGTATCTGGGCAAGACCCCAAGTTCCGCTCCGCTCCGATCCAGAGGAGA 1855  
 Db 4755 CCGAGAGCATCTGTATCTGGGCAAGACCCCAAGTTCAAGTCCGCTCCGATCCAGAGGAGA 4814  
 QY 1856 CTTGGGAGACCTGTGTGAGCCGACTACTGCGAGGCGCCTCTGTGATCCGAGTGGAGTTGG 1915  
 Db 4815 CTTGGGAGCCCTGTGTGAGCCGACTACTGCGAGGCGCCTCTGTGATCCGAGTGGAGTTGG 4874  
 QY 1916 TGAACACCCCGCTGTGTGAGTGTGTATCCAGCTTGAAGAGGAGCCCATCATCGGCG 1975  
 Db 4875 TGAACACCCCGCTGTGTGAGTGTGTATCCAGCTTGAAGAGGAGCCCATCATCGGCG 4934  
 QY 1976 CCGAGACCTTCTACCTGTGACGCGCGCCGACACCGGAGACCAAGATCGGCAAGCGCGCT 2035  
 Db 4935 CCGAGACCTTCTACCTGTGACGCGCGCCGACACCGGAGACCAAGATCGGCAAGCGCGCT 4994  
 QY 2036 ACCTGACCGACCGGCGCGCAGAGATCTGTGAGCTTGAAGAGACCAAGAGAGAGAGAG 2095  
 Db 4995 ACCTGACCGACCGGCGCGCAGAGATCTGTGAGCTTGAAGAGAGAGAGAGAGAGAGAG 5054  
 QY 2096 CCGAGCTGAGGCGCATCTGAGCTGCGCTGCGAGGACGCGGAGAGAGAGAGAGAGAG 2155  
 Db 5055 CCGAGCTGAGGCGCATCTGAGCTGCGCTGCGAGGACGCGGAGAGAGAGAGAGAGAGAG 5114  
 QY 2156 CCGAGCGGAGTACGCTTGGGCTATCTCAGGCGCGCAGCGAGAGAGAGAGAGAGAGAG 2215  
 Db 5115 CCGAGCGGAGTACGCTTGGGCTATCTCAGGCGCGCAGCGAGAGAGAGAGAGAGAGAG 5174  
 QY 2216 TGGTGAACAGATCATCTGAGCAGCTGTATCAAGAGAGAGAGAGAGAGAGAGAGAGAG 2275  
 Db 5175 TGGTGAACAGATCATCTGAGCAGCTGTATCAAGAGAGAGAGAGAGAGAGAGAGAGAG 5234  
 QY 2276 CCGCGCACAGGCGATCGGCGGCAACGAGCAGATCGACAGCTGTGTGAGCAAGGCGATCC 2335  
 Db 5235 CCGCGCACAGGCGATCGGCGGCAACGAGCAGATCGACAGCTGTGTGAGCAAGGCGATCC 5294  
 QY 2336 GCAAGTGTCTGTCTCTGAGCGATCGATGCGGATCTGTATCTACAGTA 2387  
 Db 5295 GCAAGTGTCTGTCTCTGAGCGATCTGACAGGCGCGGAGGAGCAGAGAA 5346

RESULT 12  
 AX427931  
 LOCUS 9169 bp DNA linear PAT 20-JUN-2002  
 DEFINITION Sequence 169 from Patent WO0232943.  
 ACCESSION AX427931  
 VERSION AX427931.1 GI:21538018  
 KEYWORDS synthetic construct  
 SOURCE synthetic construct  
 ORGANISM artificial sequences.  
 REFERENCE 1  
 AUTHORS Huang, Y. and Nabel, G.J.  
 TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization

JOURNAL Patent: WO 0232943-A 169 25-APR-2002;  
 GOVERNMENT OF THE UNITED STATES (US)  
 FEATURES Location/Qualifiers  
 source 1..9169  
 /organism="synthetic construct"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32630"  
 /note="plasmid pVR1012x/s containing HIV genes"  
 ORIGIN  
 Query Match 81.4%; Score 2005.2; DB 6; Length 9169;  
 Best Local Similarity 91.4%; Pred. No. 3.1e-201;  
 Matches 2187; Conservative 0; Mismatches 183; Indels 22; Gaps 5;  
 QY 14 TGGCCGAGGCGCATGAGCAGGCGCACCGCGCCACATCTGATGAGCGCAGCAATTTCA 73  
 Db 2967 TGGCCGAGGCGCATGAGCAGGCGCACCGCGCCACATCTGATGAGCGCAGCAATTTCA 3026  
 QY 74 AGGCGCCCAAGCGCATCATCAAGTGTTCAACTGCGGCAAGAGGCGCCACATCGGCCGCA 133  
 Db 3027 AGGCGCCCAAGCGCAT--CAAGTGTTCAACTGCGGCAAGAGGCGCCACCTGGGCCGCA 3083  
 QY 134 ACTGCGCGCGCCCGCGCAAGAGGCGCTGTGAGTGTGCGGCAAGAGGCGCCACAGATGA 193  
 Db 3084 ACTGCGCGCGCCCGCGCAAGAGGCGCTGTGAGTGTGCGGCAAGAGGCGCCACAGATGA 3143  
 QY 194 AGGACTGCACCGAGCGCGCAGGCGCAACTTCTTCGCGAGGACCTGTGCGCTTCCCGCAGGCA 253  
 Db 3144 AGGACTGCACCGAGCGCGCAGGCGTAA-TTTTATAGGGAAGATCTGCGCTTCCCGCAGGCA 3202  
 QY 254 AGGCGCGCGAGTTCGCCAGCGGCGAGAACCGCGCCAAAGCGCCACAGCGCGGAGCTGC 313  
 Db 3203 AGGCGCGCGAATTTCTTCAAGAGCAGACAGAGCCAAAGCGCCACAGAGAGAGAGCTTC 3262  
 QY 314 AGGTGCGCGG-----CGACAACCCCGCAGCGAGCGCGCGCCGCGCCAGGCA--- 364  
 Db 3263 AGGTTCGGGAAAGAGAGAACAACTCTCTCAGAGCAGGAGCCGATAGACAAGGAATCG 3322  
 QY 365 ---CCCTGAACTTCCCGCAGATCACCTGTGCGAGCGCCCTGTGAGCATCAAGTGG 421  
 Db 3323 TATCTTTAGCTTCCCTCAGATCACTTTTGGCAGCGACCCCTCTGTCAANTAAAGATAG 3382  
 QY 422 CGCGCCAGATCAAGGAGGCGCTCTGAGACCGCGGCGCGACACCGCTGTGAGGAGA 481  
 Db 3383 GGGCGCAGCTGAAGAGGCGCTTCTAGACACCGCGCGCGACACCGCTGTGAGGAGA 3442  
 QY 482 TGAGCTGCGCGCGCAGTGAAGCGCAAGATGATCGCGGCGCATCGCGGCTTCTCAAGG 541  
 Db 3443 TGAACCTGCGCGCGCTGAGAGCGCCAGATGATCGCGGCGCATCGCGGCTTCTCAAGG 3502  
 QY 542 TCGCGCAGTACGACCGAGATCCTGATCAGATCTGCGGCAAGAGCGCATCGGCGCGGTGC 601  
 Db 3503 TGGCGCAGTACGACCGATCCTGATCAGATCTGCGGCGCACAGGCGCATCGGCGCGGTGC 3562  
 QY 602 TGATCGCGCCCGCGCGCGTGAACATCATCGCGCGCAACATGTGACCGCGCGGTGCA 661  
 Db 3563 TGGTGGCGCCCGCGCGCGTGAACATCATCGCGCGCAACCTGTGACCGAGATCGGCTGCA 3622  
 QY 662 CCCTGAACTTCCCGCATCGCGCCCATCGAGACCGTGTGCGTGAAGCTGAAGCGCGCATGG 721  
 Db 3623 CCCTGAACTTCCCGCATCAGCGCCCATCGAGACCGTGTGCGTGAAGCTGAAGCGCGCATGG 3682  
 QY 722 ACGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGAGAGAGATCAAGGCGCTTCAAGCGCA 781  
 Db 3683 ACGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGAGAGAGATCAAGGCGCTTGGTGGAGA 3742  
 QY 782 TCTCGAGGAGATGAGAGAGGCGGCAAGATCAAGATCGGCGCGCGCGAGAGACCGCTTACA 841  
 Db 3743 TCTGACCGAGATGAGAGAGGAGGAGGCAAGATCAGCAGATCGGCGCGCGCGAGAGACCGCTTACA 3802  
 QY 842 ACACCCCGCTGTTCGCCATCAAGAGAGAGCAGCAGCAGTGGCGCGAGCTGGTGGACT 901  
 Db 3803 ACACCCCGCTGTTCGCCATCAAGAGAGAGCAGCAGCAGTGGCGCGAGCTGGTGGACT 3862





QY 273 CGAGCAGAACCGCGCCACAGCCCGCCAGCGCGAGCTGCAGGTGCGCGG-----CGA 326  
DB 3228 AGAGCAGACAGAGCCCAACAGCCCCACAGAAAGAGAGCTTCAGGTTGGGGAAGAGCAA 3287  
QY 327 CAACCCCGCAGCGCGCGCGCCGAGCGCCAGGGCA-----CCCTGAATTCGCCCA 380  
DB 3288 CAACTCCCTCTCAGAAAGCAGGAGCGCATAGACAAGGAACGTATCTTTAGCTTCCCTCA 3347  
QY 381 GATCACCCTGTGGCAGCGCCCTGCTGAGCATCAAGTGGCGCGCGAGATCAAGAGGCG 440  
DB 3348 GATCATCTTTGGCAGCGACCCCTGCTGACATAAAGATAGGGGGCCAGCTGAGGAGGC 3407  
QY 441 CTTGCTGAGACACCGCGCCGAGACACCGTGTGGAGGAGATGAACCTGCGCCCGCAAGTG 500  
DB 3408 CTTCTAGACACCGCGCCGAGACACCGTGTGGAGGAGATGAACCTGCGCGCGCGCTG 3467  
QY 501 GAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCGCATGACACAGAT 560  
DB 3468 GAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCGCATGACACAGAT 3527  
QY 561 CTTGATCGAGATCTGGGGCAAGAGGCCATCGGCACCGTGTGATCGGCCCCACCCCCGT 620  
DB 3528 CTTGATCGAGATCTGGGGCAAGAGGCCATCGGCACCGTGTGATCGGCCCCACCCCCGT 3587  
QY 621 GAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTGAACCTTCCCATCAG 680  
DB 3588 GAACATCATCGCGCGCAACATGCTGACCCAGATCGGCTGCACCCCTGAACCTTCCCATCAG 3647  
QY 681 CCCATCGACACCGTCCCTGTAAGTGAAGCCCGCATGAGCGCCCCCAAGGTGAAGCA 740  
DB 3648 CCCATCGACACCGTCCCTGTAAGTGAAGCCCGCATGAGCGCCCCCAAGGTGAAGCA 3707  
QY 741 GTGGCCCTGACGAGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGAGAA 800  
DB 3708 GTGGCCCTGACGAGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGAGAA 3767  
QY 801 GGAGGCAAGATCACCAAGATCGGCCCGAGAACCCCTCAACAACACCCCGCTGTTCCGCAT 860  
DB 3768 GGAGGCAAGATCAGCAAGATCGGCCCGAGAACCCCTCAACAACACCCCGCTGTTCCGCAT 3827  
QY 861 CAAGAAGAGACAGCACCAAGTGGCGCAAGCTGTGTGGAATTCGCGAGCTGAACAGCG 920  
DB 3828 CAAGAAGAGACAGCACCAAGTGGCGCAAGCTGTGTGGAATTCGCGAGCTGAACAGCG 3887  
QY 921 CACCCAGACTTCTGGAGGTGCAAGTGGGCTATCCCGCATCCCGCGCGCGCTGAGAGAA 980  
DB 3888 CACCCAGACTTCTGGAGGTGCAAGTGGGCTATCCCGCATCCCGCGCGCGCTGAGAGAA 3947  
QY 981 GAAGAGCGTGACCGCTTGCAGCTGGCGAGCGCCTACTTCAGCGTGCCTTGAAGCAGAA 1040  
DB 3948 GAAGAGCGTGACCGCTTGCAGCTGGCGAGCGCCTACTTCAGCGTGCCTTGAAGCAGAA 4007  
QY 1041 CTTCCGGAAGTACACCGCTTGCAGCTGGCGAGCGCCTACTTCAGCGTGCCTTGAAGCAGAA 1100  
DB 4008 CTTCCGGAAGTACACCGCTTGCAGCTGGCGAGCGCCTACTTCAGCGTGCCTTGAAGCAGAA 4067  
QY 1101 CTACAGGTACAACTGTGCTGCCCGAGGTGGAAGGGCAGCGCCAGCATCTTCCAGAGCAG 1160  
DB 4068 CTACAGGTACAACTGTGCTGCCCGAGGTGGAAGGGCAGCGCCAGCATCTTCCAGAGCAG 4127  
QY 1161 CATGACCAAGATCTGGAGCGCTTCCGCGCGCGCAACCCCGAGATCGTGTACTACCA--- 1217  
DB 4128 CATGACCAAGATCTGGAGCGCTTCCGCAAGCAGAACCCCGAGATCTGTATCTACAGTA 4187  
QY 1218 ---GGCCCCCTGTACGTGGGCGAGCGACTTGGAGATCGGCCAGCACCGCGCAAGATCGA 1274  
DB 4188 CATGACCACTGTACGTGGGCGAGCGACTTGGAGATCGGCCAGCACCGCGCAAGATCGA 4247  
QY 1275 GGAGCTGGCAAGACCTGCTGCGCTGGGGCTTACCAACCCCGAGAGAGACCCAGAA 1334  
DB 4248 GGAGCTGGCGCAGCACTGCTGCTGGGGCTTACCAACCCCGAGAGAGACCCAGAA 4307  
QY 1335 GGAGCCCCCTTCTCTGTGGATGGGCTACGAGTGCACCCCGACCAAGTGGACCGTGCAGCC 1394

DB 4308 GGAGCCCCCTTCTCTGTGGATGGGCTACGAGTGCACCCCGCAAGTGCACCGTGCAGCC 4367  
QY 1395 CATCGAGCTGCCCGCAAGAGAGAGCTGACCCGTGAACGACATCCAGAAGCTGTTGGGCAA 1454  
DB 4368 CATCTGTCTGCCCGAAGAGAGAGCTGGACCGTGAACGACATCCAGAAGCTGTTGGCAA 4427  
QY 1455 GCTGAATCTGGCGCAGCCAGATCTACCCCGCATCAAGTGGCGCGAGCTGTGCAAGCTGT 1514  
DB 4428 GCTGAATCTGGCGCAGCCAGATCTACCCCGCATCAAGTGGCGCGAGCTGTGCAAGCTGT 4487  
QY 1515 GCGCGCGCCCAAGGCCCTGACCGACATCGTGCCTCTGACCGAGGAGCGCGAGCTGGAGCT 1574  
DB 4488 GCGCGCACCAAGGCCCTGACCGAGTGGTGGCCCTGACCGAGGAGCGCGAGCTGGAGCT 4547  
QY 1575 GCGCGAAGACCGGAGATCTGCGGAGCCGTGCAAGGCTGTACTACGACCCCGCAA 1634  
DB 4548 GCGCGAAGACCGGAGATCTGGAAGGAGCGCGTGCACGCGCTGTACTACGACCCCGCAA 4607  
QY 1635 GGRCTGCTGGCGAGATCCAGAAAGCAGGCGCACGACCACTGAGTGCATCCAGATCTACCA 1694  
DB 4608 GGRCTGATCGCGAGATCCAGAAAGCAGGCGCAGGCGCAGTGCATCCAGATCTACCA 4667  
QY 1695 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGGCCAGATCGGCACCGCCACACCAA 1754  
DB 4668 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGGCCAGATCGGCAGCGCGCCACACCAA 4727  
QY 1755 CGAGTGAAGCAGCTGACCGAGCGCTGCGAAGAGATCGCCATCGAGAGCATCTGTGATCTG 1814  
DB 4728 CGAGTGAAGCAGCTGACCGAGCGCTGCGAAGAGATCGCCATCGAGAGCATCTGTGATCTG 4787  
QY 1815 GGGCAAGACCCCGCAAGTTCGCTGCCATTCGAAAGAGACCTGGGAGACCTGTGGAC 1874  
DB 4788 GGGCAAGACCCCGCAAGTTCGCTGCCATTCGAAAGAGACCTGGGAGACCTGTGGAC 4847  
QY 1875 CGACTACTGGCAGCGCCTGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGCT 1934  
DB 4848 CGAGTACTGGCAGCGCCTGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGCT 4907  
QY 1935 GAAGCTGTGTGATCCAGCTGGAGAGAGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGA 1994  
DB 4908 GAAGCTGTGTGATCCAGCTGGAGAGAGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGA 4967  
QY 1995 GCGCGCCCGCAACCGCGAGACCAAGATCGGAGCGCGCTACCTGACCGACCGCGCGCG 2054  
DB 4968 GCGCGCCCGCAACCGCGAGACCAAGCTGGGCAAGCGCGCTACCTGACCGACCGCGCGCG 5027  
QY 2055 GCAGAAAGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTGAGGCGCATCCA 2114  
DB 5028 CCAGAGGTGTGCTGCCCTGACCGACACCAACCAAGAGACCGAGCTGAGGCGCATCCA 5087  
QY 2115 GCTGCGCTGCAAGGACAGCGCAGGAGTGAACATCGTGACCGACAGCGCGAGTACGCGCT 2174  
DB 5088 CTTGCGCTTGCAGGACAGCGCGCTGAGGTGAACATCGTGACCGACAGCGCGAGTACGCGCT 5147  
QY 2175 GGGCATCATCCAGGCGCGACCGCAAGAGCGAGCGAGCTGGTGAAACCAAGATCATCGA 2234  
DB 5148 GGGCATCATCCAGGCGCGACCGCAAGAGCGAGCGAGCTGGTGAGCCAGATCATCGA 5207  
QY 2235 GCAGCTGATCAAGAAAGAGAGTGTACTGAGCTGGGTGCGCGCGCGCGCGCGCATCGG 2294  
DB 5208 GCAGCTGATCAAGAAAGAGAGTGTACTGAGCTGGGTGCGCGCGCGCGCGCGCATCGG 5267  
QY 2295 CGGCAACGAGCAGATCGACAAGCTGTGAGCAAGGCGCATCCGCAAGGTGCTTCTCTCGA 2354  
DB 5268 CGGCAACGAGCAGTGGAGCGGCTGTTGAGCGCGCGCGCGCATCCGCAAGGTGCTTCTCTCGA 5327  
QY 2355 CGGCACTGATGGCGCGCATCTGATCTACAGTA 2387  
DB 5328 CGGCACTGACAAGGCGCGCGAGGAGCAGCAGAA 5360

RESULT 14

AX427925  
LOCUS AX427925 9194 bp DNA linear PAT 20-JUN-2002  
DEFINITION Sequence 163 from Patent WO0232943.  
ACCESSION AX427925  
VERSION AX427925.1 GI:21538012  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1  
AUTHORS Huang, Y. and Nabel, G.J.  
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization  
JOURNAL Patent: WO 0232943-A 163 25-APR-2002;  
GOVERNMENT OF THE UNITED STATES (US)  
FEATURES  
source  
1. .9194  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="plasmid pVR1012x/s containing HIV genes"

ORIGIN

Query Match 81.2%; Score 2000.2; DB 6; Length 9194;  
Best Local Similarity 91.2%; Pred. No. 1e-200;  
Matches 2165; Conservative 0; Mismatches 189; Indels 19; Gaps 4;

QY 33 GGCACACGCGCAATCTGATGACGCGAGGCAATTTCAAGGGCCCAAGCGCATCAT 92  
Db GACCAACAGCGCCACCATCATGATGACGCGGCAATTTCCGCAACACGAGCGAATCGT 3048

QY 93 CAAGTGTCTCAATCGGCAAGAGGCGCAATGCGCGCAACTGCGCGGCCCGCGCAA 152  
Db GAAGTGTCTCAATCGGCAAGAGGCGCAATGCGCGGCCCACTGCGCGGCCCGCGCAA 3108

QY 153 GAAGGGCTCTGGAAGTGGCAAGAGGCGCAATGAGATGAAGATGTCACCGAGGCCA 212  
Db GAAGGGCTCTGGAAGTGGCAAGAGGCGCAATGAGATGAAGATGTCACCGAGGCCA 3168

QY 213 GGCCAACTTCTTCGCGAGGACTGCGCTTCCCGAGGGCAAGCGCGAGTTTCCCGAG 272  
Db GACTAA-TTTTATAGGAGATCTGGCTTCCACAGGAGAGGAGGAGTTTCTTC 3227

QY 273 CAGAGAGAACCGCGCAACAGCGCCACAGCGCGAGTGCAGTGGCGCG-----CGA 326  
Db AGAGAGAGAACGAGCAACAGCGCCACAGAGAGAGAGGTTTCAGGTTTGGGGAAGAGCAA 3287

QY 327 CAACCCCGCAGCGAGGCGCGCGCGAGCGCGAGCGCA-----CCCTGAACTTCCCGCA 380  
Db CAACTCCCTCTCAGAGAGAGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3347

QY 381 GATCACTCTTGGCAGCGACCGCTTGGTGGATCAAGTGGCGGCGCGAGATCAAGGAGGC 440  
Db GATCACTCTTGGCAGCGACCGCTTGGTGGATCAAGTGGCGGCGCGAGGAGGC 3407

QY 441 CCTGCTGGACACCGCGCGCGAGCACCGCTGCTGGAGGAGATGAGCTGCGCGAGTG 500  
Db CTTCTAGACACCGCGCGCGAGCACCGCTGCTGGAGGAGATGAGCTGCGCGCGCGCTG 3467

QY 501 GAAGCCCAAGATGATGCGCGGATCGCGGCTTCAAGTGGCGCGAGTACGACCCAGAT 560  
Db GAAGCCCAAGATGATGCGCGGATCGCGGCTTCAAGTGGCGCGAGTACGACCCAGAT 3527

QY 561 CTTGATCAGATCTCGCGCAAGAGGCGCATCGCGACCGTGTGTATCGCGCGCGCGCGCT 620  
Db CTTGATCAGATCTCGCGCAAGAGGCGCATCGCGACCGTGTGTATCGCGCGCGCGCGCT 3587

QY 621 GAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGCACTTGAATTTCCCGCATCAG 680  
Db GAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGCACTTGAATTTCCCGCATCAG 3687

QY 681 CCCCATCAGACCGCGCGCTGAGCTGAAGCGCGGATGAGCGCGCGCGCGCGCGCGCGCG 740

Db 3648 CCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGATGAGCGGCCCGCGGAGCA 3707  
QY 741 GTGGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCCATCTGGAGGAGATGAGAA 800  
Db 3708 GTGGCCCTGACCGAGGAGAGATCAAGGCGCTGAGAGATCTGACCGAGATGAGAA 3767  
QY 801 GGAGGGCAAGATCAACCAAGATCGCGCCCGAGAACCCCTACAAACACCCCGCTGTCGCAT 860  
Db 3768 GGAGGGCAAGATCAACCAAGATCGCGCCCGAGAACCCCTACAAACACCCCGCTGTCGCAT 3827  
QY 861 CAAGAAG 920  
Db 3828 CAAGAAG 3887  
QY 921 CACCCAGAGACTTCTGGGAGGTGAGCTGGGATCCCCCACCCCGCGGCTGAGAGAGAA 980  
Db 3888 CACCCAGAGACTTCTGGGAGGTGAGCTGGGATCCCCCACCCCGCGGCTGAGAGAGAA 3947  
QY 981 GAAGAGCTGACCGTGTGAGAGCTGGGAGCGCTACTTTCAGCGTGGCGCTGAGAGAGAA 1040  
Db 3948 GAAGAGCTGACCGTGTGAGAGCTGGGAGCGCTACTTTCAGCGTGGCGCTGAGAGAGAA 4007  
QY 1041 CTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAAGAGAGAGAGAGAGAGAG 1100  
Db 4008 CTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAAGAGAGAGAGAGAGAGAG 4067  
QY 1101 CTACAGTACAACTGTGCTGCTGAGGCTGGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1160  
Db 4068 CTACAGTACAACTGTGCTGCTGAGGCTGGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 4127  
QY 1161 CATGACCAAGATCTTGGAGCGCTTCCGCGCGCGCAACCGAGATGCTGATCTTACCA-- 1217  
Db 4128 CATGACCAAGATCTTGGAGCGCTTCCGCGCGCGCAACCGAGATGCTGATCTTACCA-- 4187  
QY 1218 ---GGCCCCCTGTACTGTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1274  
Db 4188 CATGACCAAGATCTTGGAGCGCTTCCGCGCGCGCAACCGAGATGCTGATCTTACCA-- 4247  
QY 1275 GGAGTGGCGAAGACCACTGCTGCTGGGCTTCCACACCGCGAGAGAGAGAGAGAGAGAGAG 1334  
Db 4248 GGAGTGGCGAAGACCACTGCTGCTGGGCTTCCACACCGCGAGAGAGAGAGAGAGAGAGAG 4307  
QY 1335 GGAGCCCCCTTCTGTGTGATGAGCTGACAGCTGCAACCGCGAGAGAGAGAGAGAGAGAGAG 1394  
Db 4308 GGAGCCCCCTTCTGTGTGATGAGCTGACAGCTGCAACCGCGAGAGAGAGAGAGAGAGAGAG 4367  
QY 1395 CATCGAGTGGCGGAG 1454  
Db 4368 CATCGTGTGCGCGAG 4427  
QY 1455 GCTGAACCTGGGCGGAG 1514  
Db 4428 GCTGAACCTGGGCGGAG 4487  
QY 1515 GCGGGCGGCAAGGCGCGGAG 1574  
Db 4488 GCGGGCGGCAAGGCGCGGAG 4547  
QY 1575 GCGCGAGAACCGCGAGATCTTGGCGGAGCGCGCTGCAAGAGAGAGAGAGAGAGAGAGAGAG 1634  
Db 4548 GCGCGAGAACCGCGAGATCTTGGCGGAGCGCGCTGCAAGAGAGAGAGAGAGAGAGAGAGAG 4607  
QY 1635 GGAAGTGGTGGCGGAG 1694  
Db 4608 GGAAGTGGTGGCGGAG 4667  
QY 1695 GGAGCCCTTCAAGAACCTGAG 1754  
Db 4668 GGAGCCCTTCAAGAACCTGAG 4727  
QY 1755 CGAGTGAAG 1814  
Db 4728 CGAGTGAAG 4787

QY 1815 GGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAGACCTGGGAGACCTGGTGGAC 1874  
 Db 4788 GGGCAAGACCCCAAGTTCAAGTCTCCCATCCAGAGGAGACCTGGGAGGCTGGTGGAC 4847  
 QY 1875 CGACTACTGGCGGCGCACCTGATCCCGAGTGGAGTTCTGAAACACCCGCCCTGGT 1934  
 Db 4848 CGAGTACTGGCGGCGCACCTGATCCCGAGTGGAGTTCTGAAACACCCGCCCTGGT 4907  
 QY 1935 GAAGCTGTGTTACCAAGCTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTACGTGA 1994  
 Db 4908 GAAGCTGTGTTACCAAGCTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTACGTGA 4967  
 QY 1995 CGGCGCGCGCAACCGCGAGACCAAGATCGGCAAGGCGGCTACCTGACCCGACCGGGGCGG 2054  
 Db 4968 CGGCGCGCGCAACCGCGAGACCAAGTGGCAAGGCGGCTACCTGACCCGACCGGGGCGG 5027  
 QY 2055 CGAGAGATCTGAGCTGACCGAGACCAACCAACAGAGACCGAGTGGAGGAGCATCCA 2114  
 Db 5028 CGAGAGATCTGAGCTGACCGAGACCAACCAACAGAGACCGAGTGGAGGAGCATCCA 5087  
 QY 2115 GTGBCCTCTGAGGACAGCGGCGAGGAGTGAACATCTGTGACCCAGACGACGACGCT 2174  
 Db 5088 CTTGCGCTTCGAGGACAGCGGCTGGAGGTGAACATCTGTGACCCAGACGACGACGCT 5147  
 QY 2175 GGGCATCATCCAGGCGCGGCGGACAGAGCGAGCGAGTGTGTGAACAGATCATCGA 2234  
 Db 5148 GGGCATCATCCAGGCGCGGCGGACAGAGCGAGCGAGTGTGTGAACAGATCATCGA 5207  
 QY 2235 CGAGCTGATCAAGAGGAGAGGTGATCTGAGCTGGTGGCGGCGGCGGCGGCGGCGG 2294  
 Db 5208 CGAGCTGATCAAGAGGAGAGGTGATCTGAGCTGGTGGCGGCGGCGGCGGCGGCGG 5267  
 QY 2295 CGGCAACGAGGAGATCGAAGCTGGTGGAGCAAGGAGCGGAGTGTGTGTGCTGGA 2354  
 Db 5268 CGGCAACGAGGAGATCGAAGCTGGTGGAGCGGCGGAGCGGAGTGTGTGTGCTGGA 5327  
 QY 2355 CGGCAACGAGGAGATCGAAGCTGGTGGAGCGGCGGAGCGGAGTGTGTGTGCTGGA 2387  
 Db 5328 CGGCAACGAGGAGATCGAAGCTGGTGGAGCGGCGGAGCGGAGTGTGTGTGCTGGA 5360

RESULT 15  
 AX427927 AX427927 12411 bp DNA linear PAT 20-JUN-2002  
 LOCUS Sequence 165 from Patent WO0232943.  
 DEFINITION AX427927  
 ACCESSION AX427927  
 VERSION AX427927.1 GI:21538014  
 KEYWORDS  
 SOURCE synthetic construct  
 ORGANISM synthetic construct  
 REFERENCE 1  
 AUTHORS Huang, Y. and Nabel, G.J.  
 TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for  
 JOURNAL genetic immunization  
 GOVERNMENT OF THE UNITED STATES (US)  
 FEATURES  
 1. Location/Qualifiers  
 /organism="synthetic construct"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32630"  
 /note="plasmid pVR1012x/s containing HIV genes"

ORIGIN  
 Query Match 81.2%; Score 2000.2; DB 6; Length 12411;  
 Best Local Similarity 91.2%; Pred. No. 9.6e-201;  
 Matches 2165; Conservative 0; Mismatches 189; Indels 19; Gaps 4;  
 QY 33 GGGCAAGGCGGCAACATCTGATGAGCGGCGGCAACTCCGCAACCGAGCAAGATCGT 3048  
 Db 2989 GAGCAAGGCGGCGGCAACATCTGATGAGCGGCGGCAACTCCGCAACCGAGCAAGATCGT 3048

QY 93 CAAGTGTCTCAACTGGCGCAAGGAGGCGCACATCGCCCGCAACTGCGCGCCCCCGCAAA 152  
 Db 3049 GAAGTGTCTCAACTGGCGCAAGGAGGCGCACACCGCGCGCAACTGCGCGCCCCCGCAAA 3108  
 QY 153 GAAGGCTGTGGAAGTGGCGCAAGGAGGCGCACACAGATGAAGAGTGCACCGAGCGCA 212  
 Db 3109 GAAGGCTGTGGAAGTGGCGCAAGGAGGCGCACACAGATGAAGAGTGCACCGAGCGCA 3168  
 QY 213 GGCCAAACTTTTCGCGAGGACCTGCGCTTCCCGCAGGCAAGGCGCGAGTTCCCGAG 272  
 Db 3169 GGCCTAA-TTTTTAGGGAAGATCTGCGCTTCCCGCAAGGGAAGGCGCGAGAAATTTTCTTC 3227  
 QY 273 CGAGCAGAACCGCGCCAAACAGCCCAACAGCCGCGAGTGCAGGTGCGCGG- - - - -CGA 326  
 Db 3228 AGAGCAGAACAGAGCCAAACAGCCCAACAGAGAGAGCTTCAGTTTGGGGAAGAGCA 3287  
 QY 327 CAACCCCGCAGAGGCGCGCGCCCGAGCGCCAGGCA- - - - -CCCTGAACTTCCCGCA 380  
 Db 3288 CAATCCCTCTCTGAGAGCAGGAGCCGATAGACAAGGAACTGTATCTTTAGTTTCCCTCA 3347  
 QY 381 GATCACCTCTGCGCAGCGCCCTCTGCTGAGCATCAAGGTGGCGCGCCAGATCAAGAGGC 440  
 Db 3348 GATCACTCTTTGGCAGCGACCTCTCTGCAATTAAGATAGGGGCGCCAGTGAAGGAGGC 3407  
 QY 441 CTTCTGGAACCGCGCGCCGAGCAGACACCTGCTGAGGAGATGAGCTTCCCGCGCAAGTG 500  
 Db 3408 CTTTCTAGACACCGCGCGCCGAGCAGACACCTGCTGAGGAGATGAACCTTCCCGCGCGCTG 3467  
 QY 501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAATCAAGGTGCGCCAGTACGACAGAT 560  
 Db 3468 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAATCAAGGTGCGCCAGTACGACAGAT 3527  
 QY 561 CTTGATCGAGATCTGCGGCAAGAGCCATCGGACCGTCTGAGTGCAGCCCGCCCGCGT 620  
 Db 3528 CTTGATCGAGATCTGCGGCAAGAGCCATCGGACCGTCTGAGTGCAGCCCGCCCGCGT 3587  
 QY 621 GAAATCATCGGCGCAACATGCTGACCCAGCTGGGCTGCAACCTGAACTTCCCATCAG 680  
 Db 3588 GAACATCATCGGCGCAACCTGCTGACCCAGATCGGCTGCAACCTGAACTTCCCATCAG 3647  
 QY 681 CCGCATCGAGACCGTGGCGGCAAGCTGAAAGCCCGGCAAGAGTGAAGCGCCCGCAAGTGAAGCA 740  
 Db 3648 CCGCATCGAGACCGTGGCGGCAAGCTGAAAGCCCGGCAAGAGTGAAGCGCCCGCAAGTGAAGCA 3707  
 QY 741 GTGCGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCATCTGCGAGGAGATGGAGAA 800  
 Db 3708 GTGCGCCCTGACCGAGGAGAGATCAAGGCGCTTGGTGGAGATCTGCAACCGAGATGGAGAA 3767  
 QY 801 GGAGGCAAGATCAACAGATCGGCGGCGGCAAGAGTGAAGCGCCCGCAAGTGAAGCA 860  
 Db 3768 GGAGGCAAGATCAACAGATCGGCGGCGGCAAGAGTGAAGCGCCCGCAAGTGAAGCA 3827  
 QY 861 CAAGAGAGGAGCAGCAACAAAGTGGCGCAAGTGGTGGACTTCCCGAGCTGAACAGCG 920  
 Db 3828 CAAGAGAGGAGCAGCAACAAAGTGGCGCAAGTGGTGGACTTCCCGAGCTGAACAGCG 3887  
 QY 921 CACCGAGCTTCTGGGAGTGCAGTGGGATCCCGCGGCGGCGGCGGCGGCGGCGGCGG 980  
 Db 3888 CACCGAGCTTCTGGGAGTGCAGTGGGATCCCGCGGCGGCGGCGGCGGCGGCGGCGG 3947  
 QY 981 GAAGAGCGTACCGTCTGAGCGTGGGCGACCGCTACTTTCAGCGTGGCGGCGGCGGCGG 1040  
 Db 3948 GAAGAGCGTACCGTCTGAGCGTGGGCGACCGCTACTTTCAGCGTGGCGGCGGCGGCGG 4007  
 QY 1041 CTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAACACAGAGACCCCGGCGATCCG 1100  
 Db 4008 CTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAACACAGAGACCCCGGCGATCCG 4067  
 QY 1101 CTACCAAGTACAGCTGCTGCGCGGCGGCTGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1160  
 Db 4068 CTACCAAGTACAGCTGCTGCGCGGCGGCTGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 4127



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 02:45:46 ; Search time 621 Seconds  
(without alignments)  
16849.133 Million cell updates/sec

Title: US-09-610-313-31  
Perfect score: 2463  
Sequence: 1 gtcgacgcaccatggccga.....gggtagcaccggggaattc 2463

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq 29Jan04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004as:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2463	100.0	2463	6	ABL39960 Synthetic
2	2455.4	99.7	2457	7	ACA03547 Synthetic
3	2455.4	99.7	2457	9	ADCI3265 DNA of HI
4	2442.2	99.2	2469	6	ABL39959 Synthetic
5	2436.2	98.9	2457	6	ABL39961 Synthetic
6	2430.2	98.7	2457	7	ACA03548 Synthetic
7	2430.2	98.7	2457	9	ADCI3266 DNA of HI
8	2422.6	98.4	2445	7	ACA03546 Synthetic
9	2422.6	98.4	2445	9	ADCI3264 DNA of HI
10	2415.6	98.1	3930	9	ADCI3230 DNA of HI
11	2414	98.0	3930	9	ADCI3231 DNA of HI
12	2414	98.0	3930	9	ADCI3232 DNA of HI
13	2414	98.0	5184	9	ACA03591 Synthetic
14	2414	98.0	5184	9	ADCI3279 DNA of HI
15	2383.6	96.8	3531	9	ADCI3234 DNA of HI
16	2382	96.7	3537	9	ADCI3236 DNA of HI
17	2381	96.7	5145	7	ACA03521 Synthetic
18	2381	96.7	5145	9	ADCI3233 DNA of HI
19	2371	96.3	3538	9	ADCI3235 DNA of HI
20	2322.6	94.3	3607	7	ACA03551 Synthetic
21	2322.6	94.3	3607	9	ADCI3269 DNA of HI
22	2322.6	94.3	3624	7	ACA03550 Synthetic
23	2322.6	94.3	3624	9	ADCI3268 DNA of HI

## ALIGNMENTS

### RESULT 1

ABL39960  
ID ABL39960 standard; DNA; 2463 BP.  
XX  
AC ABL39960;  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DE Synthetic construct PR975YM SEQ ID NO:31.  
XX  
KW Human immunodeficiency virus type C; antigenic HIV type C protein; immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef; immunostimulant; gene therapy; gene; ds.  
XX  
OS Human immunodeficiency virus; type C.  
OS Synthetic.  
XX  
PN WO200204493-A2.  
XX  
PD 17-JAN-2002.  
XX  
PF 05-JUL-2001; 2001WO-US021241.  
XX  
PR 05-JUL-2000; 2000US-00610313.  
XX  
PA (CHIR ) CHIRON CORP.  
XX (UYST-) UNIV STELLENBOSCH.  
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Renaburg EU;  
XX  
DR WPI; 2002-154920/20.  
XX  
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in applications including DNA immunization or generation of packaging cell lines, particularly in gene therapy.  
XX  
PS Claim 1; Fig 9; 233pp; English.  
XX  
CC The present invention describes expression cassettes comprising a polynucleotide sequence encoding a polypeptide comprising immunogenic HIV type C polypeptides. The expression cassettes comprise any of the HIV type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef (I). (I) have immunostimulant activity and can be used in gene therapy. The HIV type C polynucleotides are useful in applications including DNA immunisation, generation of packaging cell lines, and production of HIV Type C proteins. The polynucleotides are particularly useful in gene

ACA03549 Synthetic  
ADCI3267 DNA of HI  
ACA03542 Synthetic  
ACC78506 HIV p2Pol  
ACA03543 Synthetic  
ACC78507 HIV p2Pol  
ACA03541 Synthetic  
ACC78505 HIV p2Pol  
ACC78488 HIV GagPo  
ACC78489 HIV GagPo  
ACA03522 Synthetic  
ADCI3238 DNA of HI  
ACC78484 HIV GagCo  
ACC78485 HIV GagCo  
ACC78486 HIV GagCo  
ACA03584 Synthetic  
ACC78529 HIV TatRe  
ACA03592 Synthetic  
ADCI3280 DNA of HI  
ADCI3237 DNA of HI  
ACA03545 Synthetic  
ACC78509 HIV p2Pol

24 2304.4 93.6 3597 7 ACA03549  
25 2304.4 93.6 3597 9 ADCI3267  
26 2142 87.0 2466 7 ACA03542  
27 2142 87.0 2466 7 ACC78506  
28 2122.8 86.2 2472 7 ACA03543  
29 2122.8 86.2 2472 7 ACC78507  
30 2115.2 85.9 2460 7 ACA03541  
31 2115.2 85.9 2460 7 ACC78505  
32 2114.2 85.8 3564 7 ACC78488  
33 2114.2 85.8 3564 7 ACC78489  
34 2113.6 85.8 4716 7 ACA03522  
35 2113.6 85.8 4716 9 ADCI3238  
36 2109.8 85.7 3999 7 ACC78484  
37 2108.2 85.6 3999 7 ACC78485  
38 2108.2 85.6 3999 7 ACC78486  
39 2108.2 85.6 5283 7 ACA03584  
40 2108.2 85.6 5283 7 ACC78529  
41 2108 85.6 4713 7 ACA03592  
42 2108 85.6 4713 9 ADCI3280  
43 2107.6 85.6 3462 9 ADCI3237  
44 2082.2 84.5 3735 7 ACA03545  
45 2082.2 84.5 3735 7 ACC78509



QY 2041 ACCGACCGGGGGCGGAGAGATCGTGAGCCTGACCGAGACCAACCAAGACCGAG 2100  
Db 2041 ACCGACCGGGGGCGGAGAGATCGTGAGCCTGACCGAGACCAACCAAGACCGAG 2100  
QY 2101 CTGACGCCATTCAGTGGCCCTGACGACAGCGGAGCGAGTGAACATGCTGACCGAC 2160  
Db 2101 CTGACGCCATTCAGTGGCCCTGACGACAGCGGAGCGAGTGAACATGCTGACCGAC 2160  
QY 2161 AGCCAGTACGCCCTGGCCATCATCCAGGCCCGCCGACAGAGCGAGCGAGTGGTG 2220  
Db 2161 AGCCAGTACGCCCTGGCCATCATCCAGGCCCGCCGACAGAGCGAGCGAGTGGTG 2220  
QY 2221 AACCAGATCATGAGCAGCTGATCAAGAAGGAGAGTGTACTGAGCTGGGTGCCGCC 2280  
Db 2221 AACCAGATCATGAGCAGCTGATCAAGAAGGAGAGTGTACTGAGCTGGGTGCCGCC 2280  
QY 2281 CACAAGGCGATCGCGGCGCAACGAGCAGATCGCAAGCTGGTGAGCAAGGGCATCGCGAAG 2340  
Db 2281 CACAAGGCGATCGCGGCGCAACGAGCAGATCGCAAGCTGGTGAGCAAGGGCATCGCGAAG 2340  
QY 2341 GTGCTGTTCCTGACCGGCGATCGATGCGGCGATCGTGATCTACGATGACGACGACCTG 2400  
Db 2341 GTGCTGTTCCTGACCGGCGATCGATGCGGCGATCGTGATCTACGATGACGACGACCTG 2400  
QY 2401 TAGCTGGGCGAGCGCGGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2460  
Db 2401 TAGCTGGGCGAGCGCGGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2460  
QY 2461 TTC 2463  
Db 2461 TTC 2463

RESULT 2  
ACA03547  
ID ACA03547 standard; DNA; 2457 BP.  
XX ACA03547;  
AC ACA03547;  
XX  
DT 22-MAY-2003 (first entry)  
XX  
DE Synthetic DNA encoding immunogenic HIV peptide #30.  
XX  
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
KW gene therapy; packaging cell line; humoral immune response;  
KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
XX  
OS Synthetic.  
XX  
PN WO2003004657-A1.  
XX  
PD 16-JAN-2003.  
XX  
PF 05-JUL-2002; 2002WO-US021421.  
XX  
PR 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349728P.  
PR 16-JAN-2002; 2002US-0349793P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
PI Zur Megede J, Barnett SW, Lian Y;  
XX  
DR WPI; 2003-221602/21.  
XX  
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
PT polypeptides, useful as immunogenic compositions or vaccines for  
PT generating humoral or cellular immune responses against HIV in a subject,  
PT especially humans.  
XX  
PS Example 1; Fig 35; 262pp; English.

XX The invention describes a synthetic polynucleotide encoding 2 or more  
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
CC derived from different HIV subtypes. The polynucleotide is useful for  
CC immunisation. Generation of packaging cell lines, or production of HIV  
CC polypeptides. The polynucleotide and its encoded proteins are useful as  
CC immunogenic compositions or vaccines for generating humoral or cellular  
CC immune responses against HIV in a subject, or for inducing neutralising  
CC antibodies against HIV. The gene delivery vector comprising the  
CC polynucleotide is also useful for DNA immunisation of, or for generating  
CC an immune response (e.g. a humoral or cellular immune response) in, a  
CC subject such as a mammal, particularly a human. This sequence encodes a  
CC human immunodeficiency virus immunogenic peptide  
XX  
SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;  
Query Match 99.7%; Score 2455.4; DB 7; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 2.4e-296;  
Matches 2456; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GTCACGCCCATCGCGGAGCGCATGAGCGCCACGAGCGCCACCAACATCTCTGATCGAG 60  
Db 1 GTCACGCCCATCGCGGAGCGCATGAGCGCCACGAGCGCCACCAACATCTCTGATCGAG 60  
QY 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCACTCGCGCAAGGAGGCG 120  
Db 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCACTCGCGCAAGGAGGCG 120  
QY 121 CACATCGCCCGCAACTGCGCGCCCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Db 121 CACATCGCCCGCAACTGCGCGCCCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
QY 181 GGCCACACAGATGAAGAGGACTGCACCGAGCGCCCAACTTCTTCCCGAGGACCTGGCC 240  
Db 181 GGCCACACAGATGAAGAGGACTGCACCGAGCGCCCAACTTCTTCCCGAGGACCTGGCC 240  
QY 241 TTCCCGCAGGCAAGGCGCGGAGTTCCTCCAGCGAGCGAGCGCGCCACAGCCCCCACC 300  
Db 241 TTCCCGCAGGCAAGGCGCGGAGTTCCTCCAGCGAGCGAGCGCGCCACAGCCCCCACC 300  
QY 301 AGCGCGAGCTGCAGGTGCGGGCGCAACACCCCGCAGCGAGGCGCGCGCGAGCGCGAG 360  
Db 301 AGCGCGAGCTGCAGGTGCGGGCGCAACACCCCGCAGCGAGGCGCGCGCGAGCGCGAG 360  
QY 361 GGCAACCTGAACTTCCCGCAGATCACTGCGAGCGCGCGCGCGCGCGCGCGCGCGAG 420  
Db 361 GGCAACCTGAACTTCCCGCAGATCACTGCGAGCGCGCGCGCGCGCGCGCGCGCGAG 420  
QY 421 GGCGCCAGATCAAGGAGGCGCGCTGTGACACCGCGCGCGCGCGCGCGCGCGCGAG 480  
Db 421 GGCGCCAGATCAAGGAGGCGCGCTGTGACACCGCGCGCGCGCGCGCGCGCGAG 480  
QY 481 ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGAG 540  
Db 481 ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGAG 540  
QY 541 GTGCGCGAGTACGACCGACATCTGATCGAGATCTGCGGCAAGAGGCGCGCGCGCGAG 600  
Db 541 GTGCGCGAGTACGACCGACATCTGATCGAGATCTGCGGCAAGAGGCGCGCGCGCGAG 600  
QY 601 CTGATCGAG 660  
Db 601 CTGATCGAG 660  
QY 661 ACCCTGAACCTTCCCGCATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 720  
Db 661 ACCCTGAACCTTCCCGCATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 720  
QY 721 GACGAG 780  
Db 721 GACGAG 780  
QY 781 ATCTCGGAGGAGATGGAAGAGGCGCGAGATCACCAGATCGCGCGCGCGCGCGAG 840





PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.  
XX Disclosure; Fig 41; 301pp; English.  
PS The invention relates to a novel expression cassette comprising a  
XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15naseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intravenously, intranasally,  
CC subcutaneously, intradermally, transdermally, intravaginally,  
CC intrarectally, orally or intravenously. The expression cassette is useful  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC Type C related sequence of the invention.  
XX  
SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;  
Query Match 99.7%; Score 2455.4; DB 9; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 2.4e-296;  
Matches 2456; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GTCAGCCACATGCGGAGGCTATGAGCCAGCCAGCCAGCCACCATCTCTGATGCG 60  
DB 1 GTCAGCCACATGCGGAGGCTATGAGCCAGCCAGCCAGCCACCATCTCTGATGCG 60  
QY 61 GCGAGCACTTCAAGGGCCCCAAGCGCATCATCAAGTGTTCATCTCGGCAAGGGGC 120  
DB 61 GCGAGCACTTCAAGGGCCCCAAGCGCATCATCAAGTGTTCATCTCGGCAAGGGGC 120  
QY 121 CACATGCGCCGAACTGCGGCGCCCGCCGGAAGAGGGCTGCTGAAGTGGCGAAGGAG 180  
DB 121 CACATGCGCCGAACTGCGGCGCCCGCCGGAAGAGGGCTGCTGAAGTGGCGAAGGAG 180  
QY 181 GGCACACAGATGAGGACTGACAGCGCCAGCGCCAACTTCTTCCCGAGGACTCTGGCC 240  
DB 181 GGCACACAGATGAGGACTGACAGCGCCAGCGCCAACTTCTTCCCGAGGACTCTGGCC 240  
QY 241 TTCCGCCAGGCAAGGCCCGGAGTTCCCGAGCGAGCAACCGCGCCCAAGCCCCCACC 300  
DB 241 TTCCGCCAGGCAAGGCCCGGAGTTCCCGAGCGAGCAACCGCGCCCAAGCCCCCACC 300  
QY 301 AGCCGAGCTGCGAGTGGCGGCGACAAACCCCGAGCGAGGCGCGCGCGCGCGCGAG 360  
DB 301 AGCCGAGCTGCGAGTGGCGGCGACAAACCCCGAGCGAGGCGCGCGCGCGCGAG 360  
QY 361 GGCACCTTGAATTTCCCGCAGATCACCTGTGGCAGCGCCCGCTGTGAGCATCAAGTG 420  
DB 361 GGCACCTTGAATTTCCCGCAGATCACCTGTGGCAGCGCCCGCTGTGAGCATCAAGTG 420  
QY 421 GCGGCGCAGATCAAGAGGCGCTGTGACACCGCGCGCGAGCACCGTGTGGAGGAG 480  
DB 421 GCGGCGCAGATCAAGAGGCGCTGTGACACCGCGCGCGAGCACCGTGTGGAGGAG 480  
QY 481 ATGAGCTTCCCGGCAAGTGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAACAG 540  
DB 481 ATGAGCTTCCCGGCAAGTGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAACAG 540  
QY 541 GTGCGCAGATCAAGAGGCGCTGTGACACCGCGCGAGCGCCATCGGACCGTG 600  
DB 541 GTGCGCAGATCAAGAGGCGCTGTGACACCGCGCGAGCGCCATCGGACCGTG 600  
QY 601 CTGATCGGCGCCACCGCTGAACATCATCTGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
DB 601 CTGATCGGCGCCACCGCTGAACATCATCTGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
QY 661 ACCCTGAATTTCCCATAGCCCCATCGAGACCGTGTGCGGTGAAGTGAAGCCCCGGCATG 720  
DB 661 ACCCTGAATTTCCCATAGCCCCATCGAGACCGTGTGCGGTGAAGTGAAGCCCCGGCATG 720  
QY 721 GAGGCGCCCAAGTGAAGTGGCCCTGTGACCGAGGAGGATCAAGGCGCTGACCGCC 780  
DB 721 GAGGCGCCCAAGTGAAGTGGCCCTGTGACCGAGGAGGATCAAGGCGCTGACCGCC 780

DB 721 GACGGCCCCAAGGTGAAGAGTGGCCCTGTGACCGAGGAGAGATCAAGGCGCTGACCGCC 780  
QY 781 ATCTGCGAGGAGATGAGAGGAGGCAAGATCAACCAAGATCGGCGCCCGAGAAACCCCTAC 840  
DB 781 ATCTGCGAGGAGATGAGAGGAGGCAAGATCAACCAAGATCGGCGCCCGAGAAACCCCTAC 840  
QY 841 AACACCCCGCTGTGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
DB 841 AACACCCCGCTGTGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
QY 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGACAGCTGGGATCCCCCAC 960  
DB 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGACAGCTGGGATCCCCCAC 960  
QY 961 CCGCGCGCTGAAG 1020  
DB 961 CCGCGCGCTGAAG 1020  
QY 1021 AGCGTGGCCCTGAG 1080  
DB 1021 AGCGTGGCCCTGAG 1080  
QY 1081 AACAG 1140  
DB 1081 AACAG 1140  
QY 1141 CCGAGATCTTCCAG 1200  
DB 1141 CCGAGATCTTCCAG 1200  
QY 1201 GAGATCGTATCTACAGGCGCCCTGTGACGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
DB 1201 GAGATCGTATCTACAGGCGCCCTGTGACGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
QY 1261 CCGCGCAAGATCGAG 1320  
DB 1261 CCGCGCAAGATCGAG 1320  
QY 1321 AAGAGACACAG 1380  
DB 1321 AAGAGACACAG 1380  
QY 1381 TGGACCGTCAAGCCATGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
DB 1381 TGGACCGTCAAGCCATGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
QY 1441 AAGCTGGTGGGCAAGCTGAACTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
DB 1441 AAGCTGGTGGGCAAGCTGAACTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
QY 1501 CTGTGCAAGCTGTGCGCGCGCCCAAGGCGCTGACCGCATCTGTCCTGACCGAGGAG 1560  
DB 1501 CTGTGCAAGCTGTGCGCGCGCCCAAGGCGCTGACCGCATCTGTCCTGACCGAGGAG 1560  
QY 1561 GCGGAGCTGGAGTGGCGGAG 1620  
DB 1561 GCGGAGCTGGAGTGGCGGAG 1620  
QY 1621 TAGCACCCAG 1680  
DB 1621 TAGCACCCAG 1680  
QY 1681 TACAGATCTTACAG 1740  
DB 1681 TACAGATCTTACAG 1740  
QY 1741 ACCGCGCCACAG 1800  
DB 1741 ACCGCGCCACAG 1800  
QY 1801 AGCATGCTGATCTGGGCGAAG 1860  
DB 1801 AGCATGCTGATCTGGGCGAAG 1860

1861 GAGACCTGGTGGACCGACTACTGGCAGGCCAAGCTGGATCCCGAGTGGAGTTCTGTGAAC 1920  
 1861 GAGACCTGGTGGACCGACTACTGGCAGGCCAAGCTGGATCCCGAGTGGAGTTCTGTGAAC 1920  
 1921 ACCCCCCCTGGTGAAGCTGTGTACCACTGGAGAGGAGCCATCATCGGCGCCGAG 1980  
 1921 ACCCCCCCTGGTGAAGCTGTGTACCACTGGAGAGGAGCCATCATCGGCGCCGAG 1980  
 1981 ACCTTTACGTGACCGCGCCGCCAAGCCGAGACCAAGATCGGCAAGCCGCGTACGTG 2040  
 1981 ACCTTTACGTGACCGCGCCGCCAAGCCGAGACCAAGATCGGCAAGCCGCGTACGTG 2040  
 2041 ACCGACCGGCGCGGAGAGATCTGTAGCCTGACCGAGACCAACCAAGCAAGCCGAG 2100  
 2041 ACCGACCGGCGCGGAGAGATCTGTAGCCTGACCGAGACCAACCAAGCAAGCCGAG 2100  
 2101 CTCGAGGCATTCAGTGGCCCTGACGAGACCGGAGCGAGGTGAACATCGTGACCGAC 2160  
 2101 CTCGAGGCATTCAGTGGCCCTGACGAGACCGGAGCGAGGTGAACATCGTGACCGAC 2160  
 2161 AGCCAGTACGCTGGGCTATCTCAGAGCCCGAGCCGAGAGGAGGAGGAGTGGTG 2220  
 2161 AGCCAGTACGCTGGGCTATCTCAGAGCCCGAGCCGAGAGGAGGAGGAGTGGTG 2220  
 2221 AACCAATATCGAGCAGCTGTATCAAGAGGAGAGAGTGTACCTGAGTGGTGCCCGCC 2280  
 2221 AACCAATATCGAGCAGCTGTATCAAGAGGAGAGAGTGTACCTGAGTGGTGCCCGCC 2280  
 2281 CACAAGGGCATCGGCGGCAACAGCAGATCGACAAAGCTGTGAGCAAGGGCATCCGCAAG 2340  
 2281 CACAAGGGCATCGGCGGCAACAGCAGATCGACAAAGCTGTGAGCAAGGGCATCCGCAAG 2340  
 2341 GTCTGTCTGTGACCGGCTATCGATGCGCGCATCTGTATCTACGATGATCGACGACCTG 2400  
 2341 GTCTGTCTGTGACCGGCTATCGATGCGCGCATCTGTATCTACGATGATCGACGACCTG 2400  
 2401 TACGTGGGAGCGCGCGCTTACGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457  
 2401 TACGTGGGAGCGCGCGCTTACGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

RESULT 4  
 ABL39959  
 ID ABL39959 standard; DNA; 2469 BP.  
 AC ABL39959;  
 DT 15-MAY-2002 (first entry)  
 XX Synthetic construct PR975(+) SEQ ID NO:30.  
 XX Human immunodeficiency virus type C; antigenic HIV type C protein;  
 KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;  
 KW immunostimulant; gene therapy; gene; ds.  
 XX Human immunodeficiency virus; type C.  
 OS Synthetic.  
 OS  
 XX W020204493-A2.  
 XX  
 PD 17-JAN-2002.  
 XX  
 PF 05-JUL-2001; 2001WO-US021241.  
 XX  
 PR 05-JUL-2000; 2000US-00610313.  
 XX  
 PA (CHIR) CHIRON CORP.  
 PA (DIST-) UNIV STELLENBOSCH.  
 XX  
 PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;  
 XX  
 XX WPI; 2002-154920/20.

XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in  
 PT applications including DNA immunization or generation of packaging cell  
 PT lines, particularly in gene therapy.  
 XX Claim 1; Fig 8; 233pp; English.  
 XX The present invention describes expression cassettes comprising a  
 CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV  
 CC type C polypeptides. The expression cassettes comprise any of the HIV  
 CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef  
 CC (ii). (i) have immunostimulant activity and can be used in gene therapy.  
 CC The HIV type C polynucleotides are useful in applications including DNA  
 CC immunisation, generation of packaging cell lines, and production of HIV  
 CC Type C proteins. The polynucleotides are particularly useful in gene  
 CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
 CC ABL06204 to ABL06215 represent sequences used in the exemplification of  
 CC the present invention  
 XX Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;  
 SQ  
 Query Match 99.2%; Score 2442.2; DB 6; Length 2469;  
 Best Local Similarity 99.6%; Pred. No. 1.le-294;  
 Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;  
 QY 1 GTGAGCGCCACATGAGCGGAGGCGCATGAGCGGCGCACGAGCGCCCAACTCTGTATGCGAG 60  
 DB 1 GTGAGCGCCACATGAGCGGAGGCGCATGAGCGGCGCACGAGCGCCCAACTCTGTATGCGAG 60  
 QY 61 GCGAGCACTTCAAGGGGCCCCAGCGGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGCG 120  
 DB 61 GCGAGCACTTCAAGGGGCCCCAGCGGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGCG 120  
 QY 121 CACATCGCCCGCAACTGCGGCGCCCCCGGCAAGAGGGCTGTGAGAGTGGCGCAAGGAG 180  
 DB 121 CACATCGCCCGCAACTGCGGCGCCCCCGGCAAGAGGGCTGTGAGAGTGGCGCAAGGAG 180  
 QY 181 GCGCACCAGATGAAGGAGCTGCACCGAGCGCCAGGCAACTTCTTCCGCGAGGACCTTGGCC 240  
 DB 181 GCGCACCAGATGAAGGAGCTGCACCGAGCGCCAGGCAACTTCTTCCGCGAGGACCTTGGCC 240  
 QY 241 TTCCCGCGAGGCGCGCGAGTTCCTCCAGCGAGAGAAACCGCGCCCAACGCCCCAC 300  
 DB 241 TTCCCGCGAGGCGCGCGAGTTCCTCCAGCGAGAGAAACCGCGCCCAACGCCCCAC 300  
 QY 301 AGCGCGAGCTGCGAGGTCGCGCGCGCAACACCCCGCGAGCGCGCGCGCGCGCGCGAG 360  
 DB 301 AGCGCGAGCTGCGAGGTCGCGCGCGCAACACCCCGCGAGCGCGCGCGCGCGCGAG 360  
 QY 361 GCGACCTGAATTCCTCCCGAGATCACCTGTGCGAGGCGCGCGCGCGCGCGCGAGTGA 420  
 DB 361 GCGACCTGAATTCCTCCCGAGATCACCTGTGCGAGGCGCGCGCGCGCGCGAGTGA 420  
 QY 421 GCGCGCCAGATCAAGAGGCGCGCTGTGACACCGCGCGCGCGCGCGCGCGCGAGGAG 480  
 DB 421 GCGCGCCAGATCAAGAGGCGCGCTGTGACACCGCGCGCGCGCGCGCGCGAGGAG 480  
 QY 481 ATGAGCTGCGCGCAAGTGGAGCCCAAGATCATCGGCGCGCGCGCGCGCGCGAG 540  
 DB 481 ATGAGCTGCGCGCAAGTGGAGCCCAAGATCATCGGCGCGCGCGCGCGCGAG 540  
 QY 541 GTGCGCCAGTACGACCGAGATCTGATCGAGATCTGCGGCAAGAGGCGCGCGAGCGCGT 600  
 DB 541 GTGCGCCAGTACGACCGAGATCTGATCGAGATCTGCGGCAAGAGGCGCGCGAGCGCGT 600  
 QY 601 CTGATCGGCG 660  
 DB 601 CTGATCGGCG 660  
 QY 661 ACCCTGAATTCCTCCCGAGATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720  
 DB 661 ACCCTGAATTCCTCCCGAGATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720

QY	721	GACGCCCCAAGTGTGAAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC	780	1801	ATGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGGAG	1860
Db	721			1855	ACCTGGGAGACTGGTGGACCGGACTATCTGGCAGGCCACCTGGATCCCGGAGTGGAGTTTC	1914
QY	781	ATCTGCGAGGAGATGAGAGAGGGCAAGATCAACCAAGATTCGGCCCGCGAGAACCCCTTAC	840	1861	ACCTGGGAGACTGGTGGACCGGACTATCTGGCAGGCCACCTGGATCCCGGAGTGGAGTTTC	1920
Db	781			1915	GTGAACAACCCCGCCCTGCTGTAAGCTGTGTAACAGCTGGAGAGAGGCCATCATCGGC	1974
QY	841	AACACCCCGCTGTTGSCCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC	900	1921	GTGAACAACCCCGCCCTGCTGTAAGCTGTGTAACAGCTGGAGAGAGGCCATCATCGGC	1980
Db	841			1975	GCCGAGACCTTCTACGTGGACGGCGCCCAACCGCGAGACCAAGATCGGAAGCCCGGC	2034
QY	901	TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGACAGCTGGGCACTCCGCCAC	960	1981	GCCGAGACCTTCTACGTGGACGGCGCCCAACCGCGAGACCAAGATCGGAAGCCCGGC	2040
Db	901			2035	TACGTGAACGACCGCGCGCGGCAAGATCGTGAAGCTGACCGAGACCAACCAACAGAG	2094
QY	961	CCCGCGGCTGAAGAGAAAGAGCGTGACCGTGTGGAGCTGGAGCGCGCTACTTTC	1020	2041	TACGTGAACGACCGCGCGCGGCAAGATCGTGAAGCTGACCGAGACCAACCAACAGAG	2100
Db	961			2095	ACGAGCTGCAGGCCATCCAGCTGGCCCTGACGAGACGCGGAGAGAGTGAACATCGTG	2154
QY	1021	AGCGTGCCCTGGACGAGACTTCCGCAAGTACACCGCTTTCACCATCCCGAGCATCAAC	1080	2101	ACGAGCTGCAGGCCATCCAGCTGGCCCTGACGAGACGCGGAGAGAGTGAACATCGTG	2160
Db	1021			2155	ACCGAGCCAGTACGCGCTGGGCTCATCTCAGGCCCGGACCAAGAGCGAGCGAG	2214
QY	1081	AACGAGACCCCGGATCCGCTACAGTACAAAGTGTGCCCCAGGGCTGAGGGCAGC	1140	2161	ACCGAGCCAGTACGCGCTGGGCTCATCTCAGGCCCGGACCAAGAGCGAGCGAG	2220
Db	1081			2215	CTGCTGAACAGATCATCGAGCAGCTGTCAAGAGGAGAGGTGTACTGAGCTGGGTG	2274
QY	1141	CCGACATCTTCCAGAGCAGATGACCAAGATCTTGGAGCCCTTCCGCGCCCGCAACCCC	1200	2221	CTGCTGAACAGATCATCGAGCAGCTGTCAAGAGGAGAGGTGTACTGAGCTGGGTG	2280
Db	1141			2275	CCCGCCCAACAAGGGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTAGCGGCATC	2334
QY	1201	GAGATCGTGATACCA-----GGCCCCCTGTACGTGGGAGGAGGAGCTGGAGATCGGC	1254	2281	CCCGCCCAACAAGGGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTAGCGGCATC	2340
Db	1201			2335	CGCAAGGTGCTTCTCTGGACCGCATCGATCGCGCATCGTGTATCTACCACTACATGGAC	2394
QY	1255	CAGCACCGGCCAAGATCAGAGAGCTGCGAGCACCTTCTGCGTGGGGCTTCAACACC	1314	2341	CGCAAGGTGCTTCTCTGGACCGCATCGATCGCGCATCGTGTATCTACCACTACATGGAC	2400
Db	1255			2395	GACCTGTACGTGGGCGCGCGCCCTAGCATCGATTAAGAGTTTCCCGGGCTAGCACC	2454
QY	1315	CCGCAACAAGACACAGAAAGAGCCCTTCTGTTGATGGGTACGAGTGCACCCC	1374	2401	GACCTGTACGTGGGCGCGCGCCCTAGCATCGATTAAGAGTTTCCCGGGCTAGCACC	2460
Db	1315			2455	GGTGAATTC 2463	
QY	1375	GACAGTGAACGTCGAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGAC	1434	2461	GGTGAATTC 2469	
Db	1375					
QY	1381	GACAAGTGACCGTCAGCCCATCGAGCTGCCCGAAGAGAGAGTGGACCGTGAACGAC	1440			
Db	1381					
QY	1435	ATCCAGAAGCTGTGGGCAAGTGAACCTGGCCAGCCAGATCTACCCCGGCATCAAGGTG	1494			
Db	1435					
QY	1441	ATCCAGAAGCTGTGGGCAAGTGAACCTGGCCAGCCAGATCTACCCCGGCATCAAGGTG	1500			
Db	1441					
QY	1495	CGCCAGCTGTGAAGTGTGCGCGCGCAAGGCGCTGACCGACATGTGCGCCCTGACC	1554			
Db	1495					
QY	1501	CGCCAGCTGTGAAGTGTGCGCGCGCAAGGCGCTGACCGACATCGTGGCCCTGACC	1560			
Db	1501					
QY	1555	GAGGAGCGGAGTGGAGCTGGCCGAGAACCGGAGATCTTGGCGAGCCCGTGCACGCG	1614			
Db	1555					
QY	1561	GAGGAGCGGAGTGGAGCTGGCCGAGAACCGGAGATCTTGGCGAGCCCGTGCACGCG	1620			
Db	1561					
QY	1615	GTGTACTACGACCCCGCAGAGACCTGTGGCGGAGATCCAGAGCAGGCGCACGACGAG	1674			
Db	1615					
QY	1621	GTGTACTACGACCCCGCAGAGACCTGTGGCGGAGATCCAGAGCAGGCGCACGACGAG	1680			
Db	1621					
QY	1675	TGGAACCTACAGATCTACGAGGAGCCCTTCAAGACCTGAAGACCGGCAAGTACGCCAAG	1734			
Db	1675					
QY	1681	TGGAACCTACAGATCTACGAGGAGCCCTTCAAGACCTGAAGACCGGCAAGTACGCCAAG	1740			
Db	1681					
QY	1735	ATGCGCACCGCCCAACCAACGACGCTGAAGCAGCTGACCGAGCCCGTGCAGAGATCGCC	1794			
Db	1735					
QY	1741	ATGCGCACCGCCCAACCAACGACGCTGAAGCAGCTGACCGAGCCCGTGCAGAGATCGCC	1800			
Db	1741					
QY	1795	ATGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGGAG	1854			
Db	1795					

RESULT 5  
ABL39961  
ID ABL39961 standard; DNA; 2457 BP.  
XX  
AC ABL39961;  
DT 15-MAY-2002 (first entry)  
XX  
Synthetic construct PR975YMMW SEQ ID NO:32.  
XX  
Human immunodeficiency virus type C; antigenic HIV type C protein;  
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;  
KW immunostimulant; gene therapy; gene; ds.  
XX  
Human immunodeficiency virus; type C.  
OS Synthetic.  
XX  
WO200204493-A2.  
PN  
XX  
17-JAN-2002.  
PD  
XX  
05-JUL-2001; 2001WO-US021241.  
PF  
XX  
05-JUL-2000; 2000US-00610313.  
PR  
XX

PA (CHIR ) CHIRON CORP.  
 XX (UVST-) UNIV STELLENBOSCH.  
 PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;  
 XX WPI; 2002-154920/20.  
 DR  
 XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in  
 PT applications including DNA immunisation or generation of packaging cell  
 PT lines, particularly in gene therapy.  
 XX  
 XX Claim 1; Fig 10; 233pp; English.  
 PS  
 XX The present invention describes expression cassettes comprising a  
 CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV  
 CC type C polypeptides. The expression cassettes comprise any of the HIV  
 CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef  
 CC (i). (1) have immunostimulant activity and can be used in gene therapy.  
 CC The HIV type C polynucleotides are useful in applications including DNA  
 CC immunisation, generation of packaging cell lines, and production of HIV  
 CC Type C proteins. The polynucleotides are particularly useful in gene  
 CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
 CC ABL50204 to ABL506215 represent sequences used in the exemplification of  
 CC the present invention  
 XX  
 SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;  
 Query Match 98.9%; Score 2436.2; DB 6; Length 2457;  
 Best Local Similarity 99.6%; Pred. No. 5.9e-294;  
 Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY 1 GTGAGCGCCACCATGCGCGAGCGCATGAGCGAGCGCCACGAGCGCCACATCTGATGCGAG 60  
 DB 1 GTGAGCGCCACCATGCGCGAGCGCATGAGCGAGCGCCACGAGCGCCACATCTGATGCGAG 60

QY 61 CGCAGCAACTTCAAGGGGCCCCAAGCGGCATCATCAAGTCTTCAACTGCGGCAAGGCGG 120  
 DB 61 CGCAGCAACTTCAAGGGGCCCCAAGCGGCATCATCAAGTCTTCAACTGCGGCAAGGCGG 120

QY 121 CACATCGCGCGCAACTGCG 180  
 DB 121 CACATCGCGCGCAACTGCG 180

QY 181 GGCACCCAGATGAAGGACTGCAACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
 DB 181 GGCACCCAGATGAAGGACTGCAACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240

QY 241 TTCCCG 300  
 DB 241 TTCCCG 300

QY 301 AGCG 360  
 DB 301 AGCG 360

QY 361 GGCACCCAGATGAAGGACTGCAACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420  
 DB 361 GGCACCCAGATGAAGGACTGCAACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420

QY 421 GGCACCCAGATGAAGGACTGCAACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
 DB 421 GGCACCCAGATGAAGGACTGCAACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480

QY 481 ATGAGCG 540  
 DB 481 ATGAGCG 540

QY 541 GTGCG 600  
 DB 541 GTGCG 600

QY 601 CTGATCGGCG 660

DB 601 CTGATCGGCG 660

QY 661 ACCCTGGAACCTTCCCATCAGCCCCCATCGAGCGTCCCGTGAAGCTGAAGCGCGCGCATG 720

DB 661 ACCCTGGAACCTTCCCATCAGCCCCCATCGAGCGTCCCGTGAAGCTGAAGCGCGCGCATG 720

QY 721 GACG 780

DB 721 GACG 780

QY 781 ATCTGCGAGGAGATGGAAGAGGAGGCGCAAGATCACCAAGATCGGCGCGCGCGCGCGCGCG 840

DB 781 ATCTGCGAGGAGATGGAAGAGGAGGCGCAAGATCACCAAGATCGGCGCGCGCGCGCGCGCG 840

QY 841 AACACCCCGCGTGTGGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900

DB 841 AACACCCCGCGTGTGGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900

QY 901 TTCCGCGAGCTGAACAAAGCG 960

DB 901 TTCCGCGAGCTGAACAAAGCG 960

QY 961 CCG 1020

DB 961 CCG 1020

QY 1021 AGCGTGCCTTGGACGAGGAGCTTCCGCAAGTACACCGCGCTTCCACCATCCCGAGCATCAAC 1080

DB 1021 AGCGTGCCTTGGACGAGGAGCTTCCGCAAGTACACCGCGCTTCCACCATCCCGAGCATCAAC 1080

QY 1081 AACGAGAGCG 1140

DB 1081 AACGAGAGCG 1140

QY 1141 CCCAGCATTTCCAG 1200

DB 1141 CCCAGCATTTCCAG 1200

QY 1201 GAGATCGTGAATCAAGAGCG 1260

DB 1201 GAGATCGTGAATCAAGAGCG 1260

QY 1261 CG 1320

DB 1261 CG 1320

QY 1321 AAGAGAGCGAG 1380

DB 1321 AAGAGAGCGAG 1380

QY 1381 TGGACCGGTGACG 1440

DB 1381 TGGACCGGTGACG 1440

QY 1441 AAGCTGTGGCGAG 1500

DB 1441 AAGCTGTGGCGAG 1500

QY 1495 CTGTGCAAGCTGCTGCG 1554

DB 1495 CTGTGCAAGCTGCTGCG 1554

QY 1561 GCGGAGCTGGAGCTGGCGGAG 1620

DB 1561 GCGGAGCTGGAGCTGGCGGAG 1620

QY 1621 TACGACCCCGAG 1680

DB 1621 TACGACCCCGAG 1680

QY 1681 TACGAGATCTACGAG 1740

DB 1681 TACGAGATCTACGAG 1740

QY 1741 ACCGCCACACCAACGACGTCGACGAGCCGTCGAGAGATCGCCATGGAG 1800  
Db |||||  
QY 1735 ACCGCCACACCAACGACGTCGACGAGCCGTCGAGAGATCGCCATGGAG 1794  
Db |||||  
QY 1801 AGCATCGTGTACTGGGCGACAGACCCCAAGTTCCGCTGCCATCCAGAGAGAGACTGG 1860  
Db |||||  
QY 1795 AGCATCGTGTACTGGGCGACAGACCCCAAGTTCCGCTGCCATCCAGAGAGAGACTGG 1854  
Db |||||  
QY 1861 GAGACCTGGTGGACCGACTACTGCGAGCCACCTGGATCCCGAGTGGAGTTCTGTGAAC 1920  
Db |||||  
QY 1855 GAGACCTGGTGGACCGACTACTGCGAGCCACCTGGATCCCGAGTGGAGTTCTGTGAAC 1914  
Db |||||  
QY 1921 ACCGCCCCCTGGTGGAGCTGTGTACAGCTGGAGAGAGCCCATCATCGCGCCCGAG 1980  
Db |||||  
QY 1915 ACCGCCCCCTGGTGGAGCTGTGTACAGCTGGAGAGAGCCCATCATCGCGCCCGAG 1974  
Db |||||  
QY 1981 ACCTTTACTGTGGACCGCGCCCAACCGCGAGACCCAGATGCGCAAGCGCGCTACGTG 2040  
Db |||||  
QY 1975 ACCTTTACTGTGGACCGCGCCCAACCGCGAGACCCAGATGCGCAAGCGCGCTACGTG 2034  
Db |||||  
QY 2041 ACCGACCGCGCGCGAGAGATCGTAGCCTGACCGAGACCAACCAAGAGACCGAG 2100  
Db |||||  
QY 2035 ACCGACCGCGCGCGAGAGATCGTAGCCTGACCGAGACCAACCAAGAGACCGAG 2094  
Db |||||  
QY 2101 CTGACGGCCATCCAGCTGGCCCTGCAGGACAGCGCGACGAGGTGAACATGTCACCGAC 2160  
Db |||||  
QY 2095 CTGACGGCCATCCAGCTGGCCCTGCAGGACAGCGCGACGAGGTGAACATGTCACCGAC 2154  
Db |||||  
QY 2161 AGCGATGACGCTGGGATCATCATCGCGCCAGCCCGACAGAGCGAGCGAGCTGGTG 2220  
Db |||||  
QY 2155 AGCGATGACGCTGGGATCATCATCGCGCCAGCCCGACAGAGCGAGCGAGCTGGTG 2214  
Db |||||  
QY 2221 AACCAAGATCATGACGAGCTGATCAAGAGAGAGAGGTGTACCTGAGCTGGGTGCCCGCC 2280  
Db |||||  
QY 2215 AACCAAGATCATGACGAGCTGATCAAGAGAGAGAGGTGTACCTGAGCTGGGTGCCCGCC 2274  
Db |||||  
QY 2281 CACAGAGGATCGCGCGCAACGAGCAGATCGACAGCTGGTGGAGAGGCGATCGCGAAG 2340  
Db |||||  
QY 2275 CACAGAGGATCGCGCGCAACGAGCAGATCGACAGCTGGTGGAGAGGCGATCGCGAAG 2334  
Db |||||  
QY 2341 GTGCTGTTCTGGACGGCATCGATGGCGGCATCGTGATCTACCACTATGACGACCTG 2400  
Db |||||  
QY 2335 GTGCTGTTCTGGACGGCATCGATGGCGGCATCGTGATCTACCACTATGACGACCTG 2394  
Db |||||  
QY 2401 TACGTGGGACGGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2460  
Db |||||  
QY 2395 TACGTGGGACGGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2454  
Db |||||  
QY 2461 TTC 2463  
Db |||||  
QY 2455 TTC 2457

RESULT 6  
ACA03548  
ID ACA03548 standard; DNA; 2457 BP.  
XX  
AC ACA03548;  
XX  
DT 22-MAY-2003 (first entry)  
XX  
DE Synthetic DNA encoding immunogenic HIV peptide #31.  
XX  
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
KW gene therapy; packaging cell line; humoral immune response;  
KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
XX  
OS Synthetic.  
XX  
PN WO2003004657-A1.  
XX  
PD 16-JAN-2003.

XX PF 05-JUL-2002; 2002WO-US021421.  
XX PR 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-031860P.  
PR 16-JAN-2002; 2002US-0349728P.  
PR 16-JAN-2002; 2002US-0349793P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX (CHIR ) CHIRON CORP.  
XX Zur Megede J, Barnett SW, Lian Y;  
PI WPI; 2003-221602/21.  
XX  
DR New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
PT polypeptides, useful as immunogenic compositions or vaccines for  
PT generating humoral or cellular immune responses against HIV in a subject,  
PT especially humans.  
XX  
PS Example 1; Fig 36; 262pp; English.  
XX  
CC The invention describes a synthetic polynucleotide encoding 2 or more  
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
CC derived from different HIV subtypes. The polynucleotide is useful for  
CC immunisation, generation of packaging cell lines, or production of HIV  
CC polypeptides. The polynucleotide and its encoded proteins are useful as  
CC immunogenic compositions or vaccines for generating humoral or cellular  
CC immune responses against HIV in a subject, or for inducing neutralising  
CC antibodies against HIV. The gene delivery vector comprising the  
CC polynucleotide is also useful for DNA immunisation of, or for generating  
CC an immune response (e.g. a humoral or cellular immune response) in, a  
CC subject such as a mammal, particularly a human. This sequence encodes a  
CC human immunodeficiency virus immunogenic peptide  
XX  
SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

Query Match 98.7%; Score 2430.2; DB 7; Length 2457;  
Best Local Similarity 99.6%; Pred. No. 3.3e-293;  
Matches 2448; Conservative 0; Mismatches 3; Indels 6; Gaps 1;  
QY 7 GCCACCATGGCGAGCCATGAGCCAGGCCACAGGCCAACATCTGTGTCAGCGCAGC 66  
Db 1 GCCACCATGGCGAGCCATGAGCCAGGCCACAGGCCAACATCTGTGTCAGCGCAGC 60  
QY 67 AACTTCAAGGGGCCCCAAGCGCATCATCAAGTGTCTCACTGCGGCAAGAGGGCCACATC 126  
Db 61 AACTTCAAGGGGCCCCAAGCGCATCATCAAGTGTCTCACTGCGGCAAGAGGGCCACATC 120  
QY 127 GCCCGCAACTGCCCGCCCCCGCCCAAGAGGGGTGTCTGGAAGTGGCGCAAGAGGGCCAC 186  
Db 121 GCCCGCAACTGCCCGCCCCCGCCCAAGAGGGGTGTCTGGAAGTGGCGCAAGAGGGCCAC 180  
QY 187 CAGTGAAGGACTGCAACCGAGCCAGGCCCAACTTCTTCCGGAGAGACCTGGCTTCCCC 246  
Db 181 CAGTGAAGGACTGCAACCGAGCCAGGCCCAACTTCTTCCGGAGAGACCTGGCTTCCCC 240  
QY 247 CAGGGCAAGCCCGCGAGTTCCCGAGCGAGCAACCGCGCCCAACAGCCCGCCACAGCCGC 306  
Db 241 CAGGGCAAGCCCGCGAGTTCCCGAGCGAGCAACCGCGCCCAACAGCCCGCCACAGCCGC 300  
QY 307 GAGTGTGAGTGGCGGGCGAACAACCCCGAGAGAGCCCGCGCCCGAGCGCCAGGGCACC 366  
Db 301 GAGTGTGAGTGGCGGGCGAACAACCCCGAGAGAGCCCGCGCCCGAGCGCCAGGGCACC 360  
QY 367 CTGAACTTCCCCCAGATCACTTGTGGCAGCGCCCTGTGTGAGCATCAAGTGGGGCGGC 426  
Db 361 CTGAACTTCCCCCAGATCACTTGTGGCAGCGCCCTGTGTGAGCATCAAGTGGGGCGGC 420  
QY 427 CAGATCAAGGAGGCCCTGTGTGGAACAACCGCGCCGAGACACCTGTGTGAGGAGATGAGC 486  
Db 421 CAGATCAAGGAGGCCCTGTGTGGAACAACCGCGCCGAGACACCTGTGTGAGGAGATGAGC 480

487 CTGCCCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTTCATCAAGGTGGC 546  
 Db  
 481 CTGCCCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTTCATCAAGGTGGC 540  
 QY  
 547 CAGTACGACCAAGTCTGTGATCGAGATCTGCGCAAGAGGCCATCGGCAACCGTGTGATC 606  
 Db  
 541 CAGTACGACCAAGTCTGTGATCGAGATCTGCGCAAGAGGCCATCGGCAACCGTGTGATC 600  
 QY  
 607 GCGCCCAACCCCGTGAACATCTGCGCGCAACATGCTGTGACCCAGCTGGGTGACCCCTG 666  
 Db  
 601 GCGCCCAACCCCGTGAACATCTGCGCGCAACATGCTGTGACCCAGCTGGGTGACCCCTG 660  
 QY  
 667 AACTTCCCATCAGCCCATCAGACCGTGCCTGGAAGTGAAGTGAAGCCCGGATGACCGC 726  
 Db  
 661 AACTTCCCATCAGCCCATCAGACCGTGCCTGGAAGTGAAGTGAAGCCCGGATGACCGC 720  
 QY  
 727 CCCAAGTGAAGCAGTGGCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786  
 Db  
 721 CCCAAGTGAAGCAGTGGCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780  
 QY  
 787 GAGGATGAGAGAGAGGCGAGATCAACAGATCGGCCGAGAACCCCTPACAAACACC 846  
 Db  
 781 GAGGATGAGAGAGAGGCGAGATCAACAGATCGGCCGAGAACCCCTPACAAACACC 840  
 QY  
 847 CCGTGTTCGCCATCAAGAGAGAGGACAGCAGCAAGTGGCGCAAGCTGGTGAATTCGC 906  
 Db  
 841 CCGTGTTCGCCATCAAGAGAGAGGACAGCAGCAAGTGGCGCAAGCTGGTGAATTCGC 900  
 QY  
 907 GAGTGAACAGCGCAACCCAGGACTTCTGGGAGGTGACCTGGGCATCCCCACCCCGCC 966  
 Db  
 901 GAGTGAACAGCGCAACCCAGGACTTCTGGGAGGTGACCTGGGCATCCCCACCCCGCC 960  
 QY  
 967 GCGCTGAAGAGAGAGAGCGTGACCGTGTGAGCGTGGGCGACCGCTTACTTACGCGTG 1026  
 Db  
 961 GCGCTGAAGAGAGAGAGCGTGACCGTGTGAGCGTGGGCGACCGCTTACTTACGCGTG 1020  
 QY  
 1027 CCCTCGACAGAGACTTCGCAAGTACACCGCTTACCATCCCAAGATCAACACAGAG 1086  
 Db  
 1021 CCCTCGACAGAGACTTCGCAAGTACACCGCTTACCATCCCAAGATCAACACAGAG 1080  
 QY  
 1087 ACCCCCGGATCCGCTACCAAGTACAAAGTGTGCGCCAGGCTGGAAGGCGACGCCACG 1146  
 Db  
 1081 ACCCCCGGATCCGCTACCAAGTACAAAGTGTGCGCCAGGCTGGAAGGCGACGCCACG 1140  
 QY  
 1147 ATCTTCAGAGAGAGTGAACAAAGTCTGAGAGCCCTTCGCGCGCGGCAACCCGAGATC 1206  
 Db  
 1141 ATCTTCAGAGAGAGTGAACAAAGTCTGAGAGCCCTTCGCGCGCGGCAACCCGAGATC 1200  
 QY  
 1207 GTGATCTACCA-----GGCCCCCTGTAGTGGGAGAGCTGGACATCGCGCAGCAC 1260  
 Db  
 1201 GTGATCTACCAAGTACAGAGAGCTGTAGTGGGAGAGCTGGACATCGCGCAGCAC 1260  
 QY  
 1261 CGCGCCAAGATCGAGAGCTGCGCAAGCACTGCTGCGTGGGGCTTCACCAACCCCGGAC 1320  
 Db  
 1261 CGCGCCAAGATCGAGAGCTGCGCAAGCACTGCTGCGTGGGGCTTCACCAACCCCGGAC 1320  
 QY  
 1321 AAGAGCACAG 1380  
 Db  
 1321 AAGAGCACAG 1380  
 QY  
 1381 TGGACCGTGCAGCCCATCGAGCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 Db  
 1381 TGGACCGTGCAGCCCATCGAGCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 QY  
 1441 AAGCTGTGGGCAAGTGAAGTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 Db  
 1441 AAGCTGTGGGCAAGTGAAGTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 QY  
 1501 CTGTGCAAGTGTGCTGCGCGGCGCAAGGCCCTGACGACATCTGCGCCCTGACCGAGGAG 1560  
 Db  
 1501 CTGTGCAAGTGTGCTGCGCGGCGCAAGGCCCTGACGACATCTGCGCCCTGACCGAGGAG 1560  
 QY  
 1561 GCGAGCTGAGAGTGGCGCGAGAACCGCGAGATCTGCGCGAGAGAGAGAGAGAGAGAGAG 1620

1561 GCGAGCTGAGAGTGGCCGAGAACCCGCGAGATCCTGCGGAGCCCGTGCAGCGGTGTAC 1620  
 QY  
 1621 TAGCAACCCAG 1680  
 Db  
 1621 TAGCAACCCAG 1680  
 QY  
 1681 TACCAAGATCTACCAAG 1740  
 Db  
 1681 TACCAAGATCTACCAAG 1740  
 QY  
 1741 ACCGCCCAACCAACGAGCGTGAAG 1800  
 Db  
 1741 ACCGCCCAACCAACGAGCGTGAAG 1800  
 QY  
 1801 AGCATCTGTGATCTGGGCGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGAGAGAGAGAGAGAG 1860  
 Db  
 1801 AGCATCTGTGATCTGGGCGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGAGAGAGAGAGAGAG 1860  
 QY  
 1861 GAGACCTGTGTGAGCCGACTACTTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCGCTGAC 1920  
 Db  
 1861 GAGACCTGTGTGAGCCGACTACTTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCGCTGAC 1920  
 QY  
 1921 ACCCCCCCTGTGTGAGCTGTGATGAG 1980  
 Db  
 1921 ACCCCCCCTGTGTGAGCTGTGATGAG 1980  
 QY  
 1981 ACCTTCTACCTGAG 2040  
 Db  
 1981 ACCTTCTACCTGAG 2040  
 QY  
 2041 ACCGACCGGGCGCGGAG 2100  
 Db  
 2041 ACCGACCGGGCGCGGAG 2100  
 QY  
 2101 CTGACAGGACATCCAGCTGGCCCTGCGAGGACAGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160  
 Db  
 2101 CTGACAGGACATCCAGCTGGCCCTGCGAGGACAGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160  
 QY  
 2161 AGCCAGTACGCCCTGGGCGATCATCCAGGCCAGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
 Db  
 2161 AGCCAGTACGCCCTGGGCGATCATCCAGGCCAGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
 QY  
 2221 AACCAAGATCATCGAG 2280  
 Db  
 2221 AACCAAGATCATCGAG 2280  
 QY  
 2281 CACAAGGCGATCGCGCGCAACGAG 2340  
 Db  
 2281 CACAAGGCGATCGCGCGCAACGAG 2340  
 QY  
 2341 GTGCTGTCTTGGACGCGCATCGATGGCGGATCGTGTGATCTACCAAGTACAGAGAGAGAGAGAGAG 2400  
 Db  
 2341 GTGCTGTCTTGGACGCGCATCGATGGCGGATCGTGTGATCTACCAAGTACAGAGAGAGAGAGAGAG 2400  
 QY  
 2401 TACGTGGGCGAGCGCGCCCTTAGGATCGATTAAGAGCTTCCCGGGGTAGCAGCGGT 2457  
 Db  
 2401 TACGTGGGCGAGCGCGCCCTTAGGATCGATTAAGAGCTTCCCGGGGTAGCAGCGGT 2457

RESULT 7

ADCL13266  
 ID ADCL13266 standard; DNA; 2457 BP.

XX ADCL13266;

DT 18-DEC-2003 (first entry)

XX DNA of HIV construct p2Pol-opt\_C SEQ ID NO 45.

DE expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
 KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.

XX



OS	Human immunodeficiency virus.
XX	
PN	WO2003004620-A2.
XX	
PD	16-JAN-2003.
XX	
PF	05-JUL-2002; 2002WO-US021420.
XX	
PR	05-JUL-2001; 2001US-0303192P.
PR	31-AUG-2001; 2001US-031686P.
PR	16-JAN-2002; 2002US-0349871P.
XX	
PA	(CHIR ) CHIRON CORP.
PA	(UYST-) UNIV STELLENBOSCH.
XX	
PI	Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX	
DR	WPI; 2003-221593/21.
XX	
PT	New expression cassette comprising a polynucleotide sequence encoding a
PT	polypeptide including an HIV Gag Env, Int, Nef, p15NaseH, Pol, Tat,
PT	Prot, or Rev polypeptide, useful for immunization, or generating
PT	packaging cell lines.
XX	
PS	Disclosure; Fig 42; 301pp; English.
XX	
CC	The invention relates to a novel expression cassette comprising a
CC	polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC	Int, Nef, p15NaseH, Pol, Tat, Prot, or Rev polypeptides. The novel
CC	expression cassette can be used to treat HIV type C by gene therapy or
CC	used in the development of a vaccine. The gene delivery vector is
CC	administered intramuscularly, intradermally, intranasally,
CC	subcutaneously, intrarectally, transdermally, intravenously,
CC	intrarectally, orally or intravenously. The expression cassette is useful
CC	for immunisation, generating packaging cell lines and producing HIV
CC	polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC	Type C related sequence of the invention.
XX	
SQ	Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;
	Query Match 98.7%; Score 2430.2; DB 9; Length 2457;
	Best Local Similarity 99.6%; Pred. No. 3.3e-293;
	Matches 2448; Conservative 0; Mismatches 3; Indels 6; Gaps 1
Qy	7 GCACCATGGCGAGGCATGAGCAGGCGCACAGGCCCAACATCTGTATGACGAGCGCAGC 66
Dd	1 GCCACCATTGGCGAGGCCATGAGCAGGCGCACAGGCCCAACATCTGTATGACGAGCGCAGC 60
Qy	67 AACTTCAAGGGCCCCAAGGGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGGCCACATC 126
Dd	61 AACTTCAGGGCCCCAAGGGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGGCCACATC 120
Qy	127 GCCCGCAACTGCCGCGCCCCCGCGCAAGAAGGGCTGCTGGAAGTGCAGGAGGAGGGCCAC 186
Dd	121 GCCCGCAACTGCCGCGCCCCCGCGCAAGAAGGGCTGCTGGAAGTGCAGGAGGAGGGCCAC 180
Qy	187 CAGATGAAGGACTGACACGAGCGCGCAGGCCAACTTTCTTCGCGAGGACTGTGGCTTCCCC 246
Dd	181 CAGATGAAGGACTGACACGAGCGCGCAGGCCAACTTTCTTCGCGAGGACTGTGGCTTCCCC 240
Qy	247 CAGGGCAAGCGCGCGAGTTCCCAGCAGCAGACAACCGGCCACAGCCCCACACAGCGCGC 306
Dd	241 CAGGGCAAGCGCGCGAGTTCCCAGCAGCAGACAACCGGCCACAGCCCCACACAGCGCGC 300
Qy	307 GAGCTCAGGTGCGGCGGACAAACCCCGCAGCGAGCGCCGCGCGCGAGCGCCAGGGCACC 366
Dd	301 GAGCTCAGGTGCGGCGGACAAACCCCGCAGCGAGCGCCGCGCGCGAGCGCCAGGGCACC 360
Qy	367 CTGAATCTTCCCCAGATCACCTCTGTGAGCGCCCCCTTGTGTAGCATCAGGTGGCGCGC 426
Dd	361 CTGAATCTTCCCCAGATCACCTCTGTGAGCGCCCCCTTGTGTAGCATCAGGTGGCGCGC 420
Qy	427 CAGATCAAGGAGGCGCCCTGTGTACACACCGCGCGCGACACCGTGTGTGAGGAGATGAGC 486

Db 1501 CTGTGCAAGCTGCTGGCGGGCCCAAGCCCTGACCGACATCTGTGCCCTGACCGAGGAG 1560  
 Qy 1561 GCGGAGCTGGAGCTGGCGAGAACCGCGAGATCTTGGCGAGCCCTGTGCGAGGCTGTAC 1620  
 Db 1561 GCGGAGCTGGAGCTGGCGAGAACCGCGAGATCTTGGCGAGCCCTGTGCGAGGCTGTAC 1620  
 Qy 1621 TAGCACCCCAAGGACCTGTGTGCTGGAGATCCAGAGACGAGGCGCCACGACCTGGACC 1680  
 Db 1621 TAGCACCCCAAGGACCTGTGTGCTGGAGATCCAGAGACGAGGCGCCACGACCTGGACC 1680  
 Qy 1681 TACCAGATCTACAGAGCCCTTCAAGAACCTGAGACCGGCAAGTACGCGAAGATGCGC 1740  
 Db 1681 TACCAGATCTACAGAGCCCTTCAAGAACCTGAGACCGGCAAGTACGCGAAGATGCGC 1740  
 Qy 1741 ACGGCCCAACCAACGAGCTGAAGCAGCTGACCGAGCCGCTGCGAGAGATCGCCATGGAG 1800  
 Db 1741 ACGGCCCAACCAACGAGCTGAAGCAGCTGACCGAGCCGCTGCGAGAGATCGCCATGGAG 1800  
 Qy 1801 AGCATGCTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCGCCATCCAGAGGAGACCTGG 1860  
 Db 1801 AGCATGCTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCGCCATCCAGAGGAGACCTGG 1860  
 Qy 1861 GAGACTGTGTGACCGACTACTGTGAGCCGACCTGGATCCCGAGTGGAGTTCTGTAAC 1920  
 Db 1861 GAGACTGTGTGACCGACTACTGTGAGCCGACCTGGATCCCGAGTGGAGTTCTGTAAC 1920  
 Qy 1921 ACCCGCCCTGCTGTAAGCTGTGTACAGCTGTGAGAGAGGACCCATCATCGGCGCGAG 1980  
 Db 1921 ACCCGCCCTGCTGTAAGCTGTGTACAGCTGTGAGAGAGGACCCATCATCGGCGCGAG 1980  
 Qy 1981 ACTTCTAGCTGAGCGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGGCTAGCTG 2040  
 Db 1981 ACTTCTAGCTGAGCGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGGCTAGCTG 2040  
 Qy 2041 ACGGACCGGCGCGGAGAGATCTGTAGCTGACCGAGACCAACCAAGAGACCGAG 2100  
 Db 2041 ACGGACCGGCGCGGAGAGATCTGTAGCTGACCGAGACCAACCAAGAGACCGAG 2100  
 Qy 2101 CTGAGGCGCATCAGCTGCGCTGAGGACGAGCGGAGCGAGGTGAACATCGTACCGAC 2160  
 Db 2101 CTGAGGCGCATCAGCTGCGCTGAGGACGAGCGGAGCGAGGTGAACATCGTACCGAC 2160  
 Qy 2161 AGCAGTACGCTGCGGCTATCATCGAGCGCCGAGCGGCGAGAGCGAGCTGGTG 2220  
 Db 2161 AGCAGTACGCTGCGGCTATCATCGAGCGCCGAGCGGCGAGAGCGAGCTGGTG 2220  
 Qy 2221 AACGATCATCAGCAGCTGTATCAAGAGAGAGGTGTACCTGAGCTGGGTGCGCGCC 2280  
 Db 2221 AACGATCATCAGCAGCTGTATCAAGAGAGAGGTGTACCTGAGCTGGGTGCGCGCC 2280  
 Qy 2281 CACAAGGCGATCGGCGGCAACGAGCAGATCGACAAAGCTGTGTGAGCAAGGGCATCCGCAAG 2340  
 Db 2281 CACAAGGCGATCGGCGGCAACGAGCAGATCGACAAAGCTGTGTGAGCAAGGGCATCCGCAAG 2340  
 Qy 2341 GTGCTGTTCTGAGCGGATCGATCGCGGCTGCTGATCTACAGTACATGAGACCTG 2400  
 Db 2341 GTGCTGTTCTGAGCGGATCGATCGCGGCTGCTGATCTACAGTACATGAGACCTG 2400  
 Qy 2401 TAGTGGGCGAGCGGCGCTTACGATCGATTAAGAGCTTCCCGGGGCTAGCAGCCGGT 2457  
 Db 2401 TAGTGGGCGAGCGGCGCTTACGATCGATTAAGAGCTTCCCGGGGCTAGCAGCCGGT 2457

RESULT 8  
 ACA03546

ID ACA03546 standard; DNA; 2445 BP.

XX ACA03546;

XX AC

DT 22-MAY-2003 (first entry)

XX Synthetic DNA encoding immunogenic HIV peptide #29.

XX

KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
 KW Gene therapy; packaging cell line; humoral immune response;  
 KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2003004657-A1.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 05-JUL-2002; 2002WO-US021421.  
 XX  
 PR 05-JUL-2001; 2001US-0303192P.  
 PR 31-AUG-2001; 2001US-0316860P.  
 PR 16-JAN-2002; 2002US-0349728P.  
 PR 16-JAN-2002; 2002US-0349793P.  
 PR 16-JAN-2002; 2002US-0349871P.  
 XX  
 (CHIR ) CHIRON CORP.  
 PA Zur Megede J, Barnett SW, Lian Y;  
 XX  
 FI MPI; 2003-221602/21.  
 XX  
 DR New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
 PT polypeptides, useful as immunogenic compositions or vaccines for  
 PT generating humoral or cellular immune responses against HIV in a subject,  
 PT especially humans.  
 XX  
 PS Example 1; Fig 34; 262pp; English.  
 CC The invention describes a synthetic polynucleotide encoding 2 or more  
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
 CC derived from different HIV subtypes. The polynucleotide is useful for  
 CC immunisation, generation of packaging cell lines, or production of HIV  
 CC polypeptides. The polynucleotide and its encoded proteins are useful as  
 CC immunogenic compositions or vaccines for generating humoral or cellular  
 CC immune responses against HIV in a subject, or for inducing neutralising  
 CC antibodies against HIV. The gene delivery vector comprising the  
 CC polynucleotide is also useful for DNA immunisation of, or for generating  
 CC an immune response (e.g. a humoral or cellular immune response) in, a  
 CC subject such as a mammal, particularly a human. This sequence encodes a  
 CC human immunodeficiency virus immunogenic peptide  
 XX  
 SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 98.4%; Score 2422.6; DB 7; Length 2445;  
 Best Local Similarity 99.6%; Pred. No. 2.9e-292;  
 Matches 2441; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

Qy 7 GCCACCATGGCCGAGGCGCATGAGCCAGGCGCCAGCGCCCAACATCTGTATGAGGCGCAGC 66  
 Db 1 GCCACCATGGCCGAGGCGCATGAGCCAGGCGCCAGCGCCCAACATCTGTATGAGGCGCAGC 60  
 Qy 67 AACTTCAAGGGCCCCCAAGCGCATCATCAAGTGTGTTCAACTGCGGCGAAGGAGGCCCATC 126  
 Db 61 AACTTCAAGGGCCCCCAAGCGCATCATCAAGTGTGTTCAACTGCGGCGAAGGAGGCCCATC 120  
 Qy 127 GCCCGCAACTGCGCGCGCCCCCGCAAGAGAGGGGTCTGGAAGTGTGCGCAAGGAGGCCAC 186  
 Db 121 GCCCGCAACTGCGCGCGCCCCCGCAAGAGAGGGGTCTGGAAGTGTGCGCAAGGAGGCCAC 180  
 Qy 187 CAGATGAAGACTGCAACCGAGCGCGCGCAACTTCTTCCGCGAGGACCTGGCTTCCGC 246  
 Db 181 CAGATGAAGACTGCAACCGAGCGCGCGCAACTTCTTCCGCGAGGACCTGGCTTCCGC 240  
 Qy 247 CAGGCGCAAGGCGCGCGGATTTCCCGCAGCGAGCAACCGCGCCCAACAGCCCGCCAGCGCGC 306  
 Db 241 CAGGCGCAAGGCGCGCGGATTTCCCGCAGCGAGCAACCGCGCCCAACAGCCCGCCAGCGCGC 300  
 Qy 307 GAGCTGAGGCTGCGCGCGGACCAACCCCGCAGCGAGCGCGCGCGCGAGCGCGCGCGCACC 366  
 Db 301 GAGCTGAGGCTGCGCGCGGACCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCACC 360

367 CTGAATCTTCCCTCCAGATCACTCTGTGGAGCGCCCTTGTGAGCATCAAGGTGGCGGC 426  
 Db CTGAATCTTCCCTCCAGATCACTCTGTGGAGCGCCCTTGTGAGCATCAAGGTGGCGGC 420  
 QY 427 CAGATCAAGAGGCGCTCTGTGACACCGCGCGCGAGACACCGTCTGAGGAGATGAGC 486  
 Db 421 CAGATCAAGAGGCGCTCTGTGACACCGCGCGCGAGACACCGTCTGAGGAGATGAGC 480  
 QY 487 CTGCGCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAGGTGGC 546  
 Db 481 CTGCGCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAGGTGGC 540  
 QY 547 CAGTACGACCATCTCTGATCGAGATCGCGCGAGAGGCGCATCGCGCGCTTTCATCAAGGTGGC 606  
 Db 541 CAGTACGACCATCTCTGATCGAGATCGCGCGAGAGGCGCATCGCGCGCTTTCATCAAGGTGGC 600  
 QY 607 GCGCCGACCCCGTGAACATCATCGCGCGCAACATGTGACCCAGCTGGGCTGACCCCTG 666  
 Db 601 GCGCCGACCCCGTGAACATCATCGCGCGCAACATGTGACCCAGCTGGGCTGACCCCTG 660  
 QY 667 AACTTCCCATCAGCCCATCGAGACCGTGGCGTGAAGCTGAAGCCCGCGCATGGAGCGC 726  
 Db 661 AACTTCCCATCAGCCCATCGAGACCGTGGCGTGAAGCTGAAGCCCGCGCATGGAGCGC 720  
 QY 727 CCAAGGTGAAGCACTGCGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGC 786  
 Db 721 CCAAGGTGAAGCACTGCGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGC 780  
 QY 787 GAGGAGTGAAGAGAGGCGCAAGATCACCAAGATCGCGCGAGAGACCCCTCAACACC 846  
 Db 781 GAGGAGTGAAGAGAGGCGCAAGATCACCAAGATCGCGCGAGAGACCCCTCAACACC 840  
 QY 847 CCGCTGTTCCGATCAAG 906  
 Db 841 CCGCTGTTCCGATCAAG 900  
 QY 907 GAGCTGAACAGCGCACAGAGCTTCTGGAGGTGAGCTGGAGCTGGAGCTGGAGCTGGAG 966  
 Db 901 GAGCTGAACAGCGCACAGAGCTTCTGGAGGTGAGCTGGAGCTGGAGCTGGAGCTGGAG 960  
 QY 967 GGCCTGAAG 1026  
 Db 961 GGCCTGAAG 1020  
 QY 1027 CCGCTGGAGAGAGCTTCCGAGTACACCGCTTCCAGTCCAGCTCCAGCTCCAGCTCCAG 1086  
 Db 1021 CCGCTGGAGAGAGCTTCCGAGTACACCGCTTCCAGTCCAGCTCCAGCTCCAGCTCCAG 1080  
 QY 1087 ACCCCCGCATCCGCTACAGTACAACTGTGCGCCAGGGCTGGAAGGGCAGCCCGAGC 1146  
 Db 1081 ACCCCCGCATCCGCTACAGTACAACTGTGCGCCAGGGCTGGAAGGGCAGCCCGAGC 1140  
 QY 1147 ATCTTCCAGAGCAGATGACCAAGATCTCTGAGCGCTTCCGCGCCCGGCAACCCGAGATC 1206  
 Db 1141 ATCTTCCAGAGCAGATGACCAAGATCTCTGAGCGCTTCCGCGCCCGGCAACCCGAGATC 1200  
 QY 1207 GTGATCTACAGCGCCCTCTGATCGTGGCGAGCGAGCTGGAGATCGGCGCAGCCCGGCC 1266  
 Db 1201 GTGATCTACAGCGCCCTCTGATCGTGGCGAGCGAGCTGGAGATCGGCGCAGCCCGGCC 1260  
 QY 1267 AAGATCGAGGAGTGGGAG 1326  
 Db 1261 AAGATCGAGGAGTGGGAG 1320  
 QY 1327 CACCAAG 1386  
 Db 1321 CACCAAG 1374  
 QY 1387 GTGAG 1446  
 Db 1375 GTGAG 1434

1447 GTGGGCAAGCTGAATCTGGGCGAGCGAGATCTACCCGGGATCAAGGTGGCGCGAGCTGTGC 1506  
 Db 1435 GTGGGCAAGCTGAATCTGGGCGAGCGAGATCTACCCGGGATCAAGGTGGCGCGAGCTGTGC 1494  
 QY 1507 AAGCTGTCTGGCGCGCAAGGCGCTGACCGAGATCTGTGCGCCCTGACCGAGAGGCGCGAG 1566  
 Db 1495 AAGCTGTCTGGCGCGCAAGGCGCTGACCGAGATCTGTGCGCCCTGACCGAGAGGCGCGAG 1554  
 QY 1567 CTGAGAGTGGCGGAGAAACCGCGAGATCTCTGGCGAGCGCGTGTGACCGCGCTGTACTACGAC 1626  
 Db 1555 CTGAGAGTGGCGGAGAAACCGCGAGATCTCTGGCGAGCGCGTGTGACCGCGCTGTACTACGAC 1614  
 QY 1627 CCCAGCAAGGACCTGTGGCGCGAGATCCAGAGCAGGCGCACACACAGTGGAGCTTACCAG 1686  
 Db 1615 CCCAGCAAGGACCTGTGGCGCGAGATCCAGAGCAGGCGCACACACAGTGGAGCTTACCAG 1674  
 QY 1687 ATCTACAGAGAGCCCTTCAAGAACCTGAGACCGGCGAGTACCGCAAGATGCGCAAGTGGCGC 1746  
 Db 1675 ATCTACAGAGAGCCCTTCAAGAACCTGAGACCGGCGAGTACCGCAAGATGCGCAAGTGGCGC 1734  
 QY 1747 CACACCAACGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAGAGCATC 1806  
 Db 1735 CACACCAACGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAGAGCATC 1794  
 QY 1807 GTGATCTGGGCGAAGACCCCGAGTTCGCGCTGCCATCCAGAGAGAGAGAGAGAGAGAGAG 1866  
 Db 1795 GTGATCTGGGCGAAGACCCCGAGTTCGCGCTGCCATCCAGAGAGAGAGAGAGAGAGAGAG 1854  
 QY 1867 TGGTGGACCGGACTACTGGCGAGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCG 1926  
 Db 1855 TGGTGGACCGGACTACTGGCGAGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCG 1914  
 QY 1927 CCGCTGTGAAGCTGTGGTACAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1986  
 Db 1915 CCGCTGTGAAGCTGTGGTACAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1974  
 QY 1987 TACGTGAAGCGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGAACCGGAC 2046  
 Db 1975 TACGTGAAGCGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGAACCGGAC 2034  
 QY 2047 CCGGCGCGGCGAAGAGATCGTGAAGCTGAGCGAGACCAACCAAGAGAGAGAGAGAGAGAG 2106  
 Db 2035 CCGGCGCGGCGAAGAGATCGTGAAGCTGAGCGAGACCAACCAAGAGAGAGAGAGAGAGAG 2094  
 QY 2107 GCCATCCAGCTGGCGCTTGCAGGAGCAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2166  
 Db 2095 GCCATCCAGCTGGCGCTTGCAGGAGCAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2154  
 QY 2167 TACGCGCTTGGCGATCATCCAGGCGCGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2226  
 Db 2155 TACGCGCTTGGCGATCATCCAGGCGCGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2214  
 QY 2227 ATCATCGAGAGCTGATCAAG 2286  
 Db 2215 ATCATCGAGAGCTGATCAAG 2274  
 QY 2287 GGCATCGGCGCGCAACGAGCAGATCGCAAGCTGTGAGCAAGGGGCTCCGCAAGGTGCTG 2346  
 Db 2275 GGCATCGGCGCGCAACGAGCAGATCGCAAGCTGTGAGCAAGGGGCTCCGCAAGGTGCTG 2334  
 QY 2347 TTCCTGAGAGGATCGATGGCGGATCGTATCTACAGTACATGAGAGAGAGAGAGAGAGAG 2406  
 Db 2335 TTCCTGAGAGGATCGATGGCGGATCGTATCTACAGTACATGAGAGAGAGAGAGAGAGAG 2394  
 QY 2407 GGCAGCGCGCGCTTGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2457  
 Db 2395 GGCAGCGCGCGCTTGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2445

RESULT 9  
 ADC13264  
 ID AD13264 standard; DNA; 2445 BP.  
 XX





Db 1727 AGGCGCGGAGTTCCACGAGGAGAACCGCGCCAAACAGCCCCACCGCGGAGCTGC 1786  
QY 314 AGGTGCGCGGAGCAAAACCCCGGAGGAGCGCGCGCGCGAGCGCCAGGGCAACCTGAACT 373  
Db 1787 AGGTGCGCGGAGCAAAACCCCGGAGGAGCGCGCGCGCGAGCGCCAGGGCAACCTGAACT 1846  
QY 374 TCCGCCAGATCACCTGTGGCAGCGCCCTCGTGAGCATCAAGTGGGCGGCCAGATCA 433  
Db 1847 TCCGCCAGATCACCTGTGGCAGCGCCCTCGTGAGCATCAAGTGGGCGGCCAGATCA 1906  
QY 434 AGAGGCGCTGTGAGCAACCGGCGCGCGAGCAACCGTGTGAGGAGATGAGCTTCCCG 493  
Db 1907 AGAGGCGCTGTGAGCAACCGGCGCGCGAGCAACCGTGTGAGGAGATGAGCTTCCCG 1966  
QY 494 GCAAGTGGAGCCCAAGATGATCGGCGGATCGGCGGCTTCAAGGTGGCGCGCGAGTACG 553  
Db 1967 GCAAGTGGAGCCCAAGATGATCGGCGGATCGGCGGCTTCAAGGTGGCGCGCGAGTACG 2026  
QY 554 ACCAGATCCGTGAGATCTGGGCAAGAGGCCATCGGCGGCTTCAAGGTGGCGCGCGAGTACG 613  
Db 2027 ACCAGATCCGTGAGATCTGGGCAAGAGGCCATCGGCGGCTTCAAGGTGGCGCGCGAGTACG 2086  
QY 614 CCCCCTGGAATCATCTGGGCGCAACATGCTGACCCAGCTGGGCTGCAACCTGAACTTCC 673  
Db 2087 CCCCCTGGAATCATCTGGGCGCAACATGCTGACCCAGCTGGGCTGCAACCTGAACTTCC 2146  
QY 674 CCATAGCCCCATCGAGACCGTCCCGTGAAGTGAAGCCCGGCGATGAGCGGCCCAAG 733  
Db 2147 CCATAGCCCCATCGAGACCGTCCCGTGAAGTGAAGCCCGGCGATGAGCGGCCCAAG 2206  
QY 734 TGAAGCAGTGCCTGACCGAGGAGAGATCAAGGCCCTGACCGGCTGCAACCTGAGGAGA 793  
Db 2207 TGAAGCAGTGCCTGACCGAGGAGAGATCAAGGCCCTGACCGGCTGCAACCTGAGGAGA 2266  
QY 794 TGGAGAGGAGGCAAGATCAACAAAGATCGGCGCGCGAGAACCCCTTCAACACCCCGTGT 853  
Db 2267 TGGAGAGGAGGCAAGATCAACAAAGATCGGCGCGCGAGAACCCCTTCAACACCCCGTGT 2326  
QY 854 TCGCCATCAAGAGAGGAGCAGCAACCAAGTGGCGGAGCTGAGTCCCGGAGCTGA 913  
Db 2327 TCGCCATCAAGAGAGGAGCAGCAACCAAGTGGCGGAGCTGAGTCCCGGAGCTGA 2386  
QY 914 ACAAGCGCACCCAGGACTTCTGGAGGTGCAAGTGGGCAATCCCGCACCCCGCGGCTGA 973  
Db 2387 ACAAGCGCACCCAGGACTTCTGGAGGTGCAAGTGGGCAATCCCGCACCCCGCGGCTGA 2446  
QY 974 AGAAGAGAGAGCGTGACCGTGTGACGTGGCGGAGCGCTACTTCAGCGTGCCTGG 1033  
Db 2447 AGAAGAGAGAGCGTGACCGTGTGACGTGGCGGAGCGCTACTTCAGCGTGCCTGG 2506  
QY 1034 ACGAGGACTTCCGCAAGTACACCGCTTCAACATCCCGCAGCATCAACAGAGACCCCG 1093  
Db 2507 ACGAGGACTTCCGCAAGTACACCGCTTCAACATCCCGCAGCATCAACAGAGACCCCG 2566  
QY 1094 GCATCCGCTACAGTACAACTGTGCTGCGCCAGGCTGGAGGGCAGCGCCAGCATCTTC 1153  
Db 2567 GCATCCGCTACAGTACAACTGTGCTGCGCCAGGCTGGAGGGCAGCGCCAGCATCTTC 2626  
QY 1154 AGAGCAGCATGACCAAGATCTCGAGCGCTTCCGCGCGCGCAACCCCGGAGATCGTGTCT 1213  
Db 2627 AGAGCAGCATGACCAAGATCTCGAGCGCTTCCGCGCGCGCAACCCCGGAGATCGTGTCT 2686  
QY 1214 ACCAGGCGCTGTACGTGGGAGGAGCACTGGAGATCGGCGCAGCAACCGCGCAGATCG 1273  
Db 2687 ACCAGGCGCTGTACGTGGGAGGAGCACTGGAGATCGGCGCAGCAACCGCGCAGATCG 2746  
QY 1274 AGGAGCTGGCAAGCACTGTGCGCTGGGGTTCACCAACCCCGGAGCAAGAGCACCAGA 1333  
Db 2747 AGGAGCTGGCAAGCACTGTGCGCTGGGGTTCACCAACCCCGGAGCAAGAGCACCAGA 2806  
QY 1334 AGGAGCGCCCTTCTGTGGATGGGGCTACGAGCTGCAACCCCGCAAGTGGACCGTGAGC 1393

Db 2807 AGGAGCGCCCTTCTGTGCCAT-----CGAGCTGCAACCCCGCAAGTGGACCGTGCAGC 2860  
QY 1394 CCATCGAGCTGCCGAGAGAGAGTGGAACCGTGAAACGACATCCAGAGCTGTGGGCA 1453  
Db 2861 CCATCGAGCTGCCGAGAGAGAGTGGAACCGTGAAACGACATCCAGAGCTGTGGGCA 2920  
QY 1454 AGCTGAACTGGGCGACCGAGATCTACCCCGGATCAAGGTGGCGCAGCTGTGCAAGCTGC 1513  
Db 2921 AGCTGAACTGGGCGACCGAGATCTACCCCGGATCAAGGTGGCGCAGCTGTGCAAGCTGC 2980  
QY 1514 TGGCGCGCGCAAGGCGCTGACCGACATCGTGCCCTGACCGAGAGGCGCGAGCTGGAGC 1573  
Db 2981 TGGCGCGCGCAAGGCGCTGACCGACATCGTGCCCTGACCGAGGAGGCGCGAGCTGGAGC 3040  
QY 1574 TGGCGGAGAACCGGAGATCTTCGGGAGCGCGTGCACCGCTGTACTACGACCCCGAGCA 1633  
Db 3041 TGGCGGAGAACCGGAGATCTTCGGGAGCGCGTGCACCGCTGTACTACGACCCCGAGCA 3100  
QY 1634 AGGACCTTGTGGCGCGAGATCCAGAAGCAGGGCCACGACCAAGTGGACCTTACAGATCTACC 1693  
Db 3101 AGGACCTTGTGGCGCGAGATCCAGAAGCAGGGCCACGACCAAGTGGACCTTACAGATCTACC 3160  
QY 1694 AGGAGCGCTTCAAGAACCTGAAGACCGGAGTACGCCAAGATCGCACCGCCACCAACCA 1753  
Db 3161 AGGAGCGCTTCAAGAACCTGAAGACCGGAGTACGCCAAGATCGCACCGCCACCAACCA 3220  
QY 1754 ACGACGTGAAGCAGCTGACCGAGCGCGTGCAGAAGATCGCCATGGAGAGCATCGTGTCT 1813  
Db 3221 ACGACGTGAAGCAGCTGACCGAGCGCGTGCAGAAGATCGCCATGGAGAGCATCGTGTCT 3280  
QY 1814 GGGCGAAGACCCCAAGTTCGGCTGCGCATCCAGAGGAGAGCTGGGAGACCTGTGTGA 1873  
Db 3281 GGGCGAAGACCCCAAGTTCGGCTGCGCATCCAGAAGGAGAGCTGGGAGACCTGTGTGA 3340  
QY 1874 CCGACTACTGCGAGCGCACCTTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGG 1933  
Db 3341 CCGACTACTGCGAGCGCACCTTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGG 3400  
QY 1934 TGAAGCTGTGTAACAGCTGGAAGAGAGCCCATCATCGCGCGCGAGACCTTCTACGTGG 1993  
Db 3401 TGAAGCTGTGTAACAGCTGGAAGAGAGCCCATCATCGCGCGCGAGACCTTCTACGTGG 3460  
QY 1994 ACGGCGCGCCAAACCGGAGACCAAGATCGGCAAGGCGGCTACGTGACCGACCGAGCGGCGCC 2053  
Db 3461 ACGGCGCGCCAAACCGGAGACCAAGATCGGCAAGGCGGCTACGTGACCGACCGAGCGGCGCC 3520  
QY 2054 GGCAGAGATCGTGAGCTGACCGAGACCAACCGAGAGACCGAGCTGCAGGCGCATCC 2113  
Db 3521 GGCAGAGATCGTGAGCTGACCGAGACCAACCGAGAGACCGAGCTGCAGGCGCATCC 3580  
QY 2114 AGCTGCGCTCGAGGACAGCGGAGTGAACATCGTGACCGGACAGCCAGTACCGCC 2173  
Db 3581 AGCTGCGCTCGAGGACAGCGGAGTGAACATCGTGACCGGACAGCCAGTACCGCC 3640  
QY 2174 TGGGCAATCATCCAGCGCCAGCGCGCAAGAGGAGAGCGAGCTGGTGAACCGAGTCACTCG 2233  
Db 3641 TGGGCAATCATCCAGCGCCAGCGCGCAAGAGGAGAGCTGGTGAACCGAGTCACTCG 3700  
QY 2234 AGCAGCTGATCAAGAGGAGAGGTGTACTCTGAGTGGGTGCGCCCGCACAGGCGCATCG 2293  
Db 3701 AGCAGCTGATCAAGAGGAGAGGTGTACTCTGAGTGGGTGCGCCCGCACAGGCGCATCG 3760  
QY 2294 GGGGCAACCGAGCAGATCGCAAGCTGGTGAAGCAAGGCAATCCGCAAGTGTGTCTCTGG 2353  
Db 3761 GGGGCAACCGAGCAGATCGCAAGCTGGTGAAGCAAGGCAATCCGCAAGTGTGTCTCTGG 3820  
QY 2354 ACGGCAATCGATGGCGCATCGTGATCTACGATACGACGACCTGTACGTGGGCGAGCG 2413  
Db 3821 ACGGCAATCGATGGCGCATCGTGATCTACGATACGACGACCTGTACGTGGGCGAGCG 3880  
QY 2414 GCGGCGCTTAGGATCGATTAAAGCTTCCCGGGCTTAGCACCGGT 2457  
Db 3881 GCGGCGCTTAGGATCGATTAAAGCTTCCCGGGCTTAGCACCGGT 3924





1334 AGAGCCCGCTTCTGTGGTGGCTACGAGCTGCACCCCGACAAAGTGGACCGTGCAGC 1393  
Db |||||  
2807 AGAGCCCGCTTCTGTGGTGGCTACGAGCTGCACCCCGACAAAGTGGACCGTGCAGC 2860  
Db |||||  
1394 CCATCAGCTGCCCGAGAGAGAGTGCACCGTGAACGACATCCAGAGCTGTGTGGCA 1453  
Db |||||  
2861 CCATCAGCTGCCCGAGAGAGAGTGCACCGTGAACGACATCCAGAGCTGTGTGGCA 2920  
Db |||||  
1454 AGCTGAATGGGCGACCGAGATCTACCCCGGATCAAGTGGCCAGCTGTCAAGCTGC 1513  
Db |||||  
2921 AGCTGAATGGGCGACCGAGATCTACCCCGGATCAAGTGGCCAGCTGTCAAGCTGC 2980  
Db |||||  
1514 TCGCGCGCCAAAGGCGCTGACCGATCGTGCCTGACCGAGAGGCGGAGCTGGAGC 1573  
Db |||||  
2981 TCGCGCGCCAAAGGCGCTGACCGATCGTGCCTGACCGAGAGGCGGAGCTGGAGC 3040  
Db |||||  
1574 TCGCGCGAGAACCGGAGATCTCGGAGCGCCGCTGACCGGCTGTACTACGACCCAGCA 1633  
Db |||||  
3041 TCGCGCGAGAACCGGAGATCTCGGAGCGCCGCTGACCGGCTGTACTACGACCCAGCA 3100  
Db |||||  
1634 AGACCTGTGTGGCGGAGATCCAGAGCAGGCGCCACGACAGTGGACCTACAGATCTACC 1693  
Db |||||  
3101 AGACCTGTGTGGCGGAGATCCAGAGCAGGCGCCACGACAGTGGACCTACAGATCTACC 3160  
Db |||||  
1694 AGAGCCCTTCAAGAACCTGAACCGGCAAGTACCGCAAGATGCGCACCGGCCACCA 1753  
Db |||||  
3161 AGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACCGCAAGATGCGCACCGGCCACCA 3220  
Db |||||  
1754 ACAGCTGAAGAGCTGACCGAGCGCGCTGACAGAGATCGCCATGGAGAGCTGTGATCT 1813  
Db |||||  
3221 ACAGCTGAAGAGCTGACCGAGCGCGCTGACAGAGATCGCCATGGAGAGCTGTGATCT 3280  
Db |||||  
1814 GGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGAGGAGCTGGAGACCTGTGTGA 1873  
Db |||||  
3281 GGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGAGGAGCTGGAGACCTGTGTGA 3340  
Db |||||  
1874 CCGACTACTGCGAGCGCACCTGTGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGG 1933  
Db |||||  
3341 CCGACTACTGCGAGCGCACCTGTGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGG 3400  
Db |||||  
1934 TGAAGCTGTGTACGAGTGGAGAGGAGCGCATCTACGCGCGGAGACCTTACGTGG 1993  
Db |||||  
3401 TGAAGCTGTGTACGAGTGGAGAGGAGCGCATCTACGCGCGGAGACCTTACGTGG 3460  
Db |||||  
1994 ACAGCGCGCGCAACCGGAGACCAAGATCGGAGCGCGCTACGTGACCGACCGGGGCC 2053  
Db |||||  
3461 ACAGCGCGCGCAACCGGAGACCAAGATCGGAGCGCGCTACGTGACCGACCGGGGCC 3520  
Db |||||  
2054 GGCAGAGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTGACGCGCATCC 2113  
Db |||||  
3521 GGCAGAGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTGACGCGCATCC 3580  
Db |||||  
2114 AGCTGGCCCTGACGAGCGGAGCGAGTGTGACATCGTGCAGCAGCGAGTACGCC 2173  
Db |||||  
3581 AGCTGGCCCTGACGAGCGGAGCGAGTGTGACATCGTGCAGCAGCGAGTACGCC 3640  
Db |||||  
2174 TGGGCATCATCCAGGCGCGAGCGGAGCGGAGCGAGCTGTGTGAACAGATCATCG 2233  
Db |||||  
3641 TGGGCATCATCCAGGCGCGAGCGGAGCGGAGCGAGCTGTGTGAACAGATCATCG 3700  
Db |||||  
2234 AGCAGCTGATCAAGAGGAGAGTGTACTGTGCTGGTGGTCCCGCCCAAGGGCATCG 2293  
Db |||||  
3701 AGCAGCTGATCAAGAGGAGAGTGTACTGTGCTGGTGGTCCCGCCCAAGGGCATCG 3760  
Db |||||  
2294 GCGCAACGAGCAGATCGACAAGCTGTGTGAACAAAGGGCATCCGCAAGGTGCTGTCTCG 2353  
Db |||||  
3761 GCGCAACGAGCAGATCGACAAGCTGTGTGAACAAAGGGCATCCGCAAGGTGCTGTCTCG 3820  
Db |||||  
2354 ACAGCATGATGCGGCGCATCTGTGATCTACAGTACATGGAGACCTGTACGTGGCAGCG 2413  
Db |||||  
3821 ACAGCATGATGCGGCGCATCTGTGATCTACAGTACATGGAGACCTGTACGTGGCAGCG 3880  
Db |||||

2414 GCGGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457  
Db |||||  
3881 GCGGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 3924  
Db |||||  
RESULT 12  
ADCI3232  
ID ADCI3232 standard; DNA; 3930 BP.  
XX  
AC ADCI3232;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE DNA of HIV construct GagComplPolmutIna\_C SEQ ID NO 11.  
XX  
KW expression cassette; HIV Gag; Env; Int; Nef; p15NaseH; Pol; Tat; Prot;  
XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
XX Human immunodeficiency virus.  
OS  
XX WO2003004620-A2.  
PN  
XX 16-JAN-2003.  
PD  
XX 05-JUL-2002; 2002WO-US021420.  
PF  
XX 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316850P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX  
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
XX WPI; 2003-221593/21.  
DR  
XX New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15NaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.  
XX  
PS Disclosure; Fig 8; 301pp; English.  
XX  
CC The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15NaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intracutaneously, intranasally,  
CC subcutaneously, intradermally, transdermally, intravaginally,  
CC intrarectally, orally or intravenously. The expression cassette is useful  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC Type C related sequence of the invention.  
XX  
SQ Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;  
Query Match 98.0%; Score 2414; DB 9; Length 3930;  
Best Local Similarity 99.5%; Pred. No. 3.1e-291;  
Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;  
QY 14 TGGCGGAGCCATGAGCCAGGCGCCACGCGCCACATCTGATGAGCGCAACTTCA 73  
Db 1487 TGGCGGAGCCATGAGCCAGGCGCCACGCGCCACATCTGATGAGCGCAACTTCA 1546  
QY 74 AGGGCCCAAGCGCATCATAGTCTTCAACTGGCGCAAGGGGCCACATCGCCGCA 133  
Db 1547 AGGGCCCAAGCGCATCATAGTCTTCAACTGGCGCAAGGGGCCACATCGCCGCA 1606  
QY 134 ACTGCGCGCCCGCCCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGCCATCA 193  
Db 1607 ACTGCGCGCCCGCCCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGCCATCA 1666

194 AGGACTGCGCAGCGCCAGCCAACTTCTTCGCGAGGACCTGGGCTTCCGCCAGGGCA 253 QY  
1667 AGGACTGCGCAGCGCCAGCCAACTTCTTCGCGAGGACCTGGGCTTCCGCCAGGGCA 1726 Db  
254 AGGCGCGGAGTTCGCCAGCGAGCAGAACCGCGCCACAGCCGCCACCGAGCGCTGC 313 QY  
1727 AGGCGCGGAGTTCGCCAGCGAGCAGAACCGCGCCACAGCCGCCACCGAGCGCTGC 1786 Db  
314 AGGTGCGCGGCGCAAAACCCCGCGAGCGAGGCGCGCGCGCGAGCGCGAGCGCACTGAACT 373 QY  
1787 AGGTGCGCGGCGCAAAACCCCGCGAGCGAGGCGCGCGCGCGAGCGCGAGCGCACTGAACT 1846 Db  
374 TCCCGCGAGATCACCTGTGGCAGCGCCCGCTGTGTGAGCATCAAGGTGGCGCGCGAGATCA 433 QY  
1847 TCCCGCGAGATCACCTGTGGCAGCGCCCGCTGTGTGAGCATCAAGGTGGCGCGCGAGATCA 1906 Db  
434 AGGAGGCGCTCTGGACACCGCGCGCGAGCAGACACCGTGTGTGGAGGAGATGAGCGTGGCCG 493 QY  
1907 AGGAGGCGCTCTGGCAGACCGCGCGCGAGCAGACACCGTGTGTGGAGGAGATGAGCGTGGCCG 1966 Db  
494 GCAAGTGAAGCGCCAGAGTATCGGCGGCATCGGCGGCTTCATCAAGGTGGCGCGAGTACG 553 QY  
1967 GCAAGTGAAGCGCCAGAGTATCGGCGGCATCGGCGGCTTCATCAAGGTGGCGCGAGTACG 2026 Db  
554 ACCAGATCCTGATCGAGATCTGCGCGCAAGAGGCGCATCGGCACCGTGTGATCGGCCCA 613 QY  
2027 ACCAGATCCTGATCGAGATCTGCGCGCAAGAGGCGCATCGGCACCGTGTGATCGGCCCA 2086 Db  
614 CCCCCTGAAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGCACCTGAACTTCC 673 QY  
2087 CCCCCTGAAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGCACCTGAACTTCC 2146 Db  
674 CCATCAGCGCCCATCGAGACCGTGCCTGAAAGCTGAAGCCCGGATGAGCGGCCCAAGG 733 QY  
2147 CCATCAGCGCCCATCGAGACCGTGCCTGAAAGCTGAAGCCCGGATGAGCGGCCCAAGG 2206 Db  
734 TGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793 QY  
2207 TGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266 Db  
794 TGGAGAGGAGGCGAAGATCAACAAAGATCGGCGCGAGAACCCCTTACAAACCCCGCTGT 853 QY  
2267 TGGAGAGGAGGCGAAGATCAACAAAGATCGGCGCGAGAACCCCTTACAAACCCCGCTGT 2326 Db  
854 TCGCCATCAAGAGAGAGAGCAGCAACAAAGTGGCGCAAGCTGGTGACTTCCGCGAGCTGA 913 QY  
2327 TCGCCATCAAGAGAGAGAGCAGCAACAAAGTGGCGCAAGCTGGTGACTTCCGCGAGCTGA 2386 Db  
914 ACAAGGCGACCCAGGACTTCTGGAGGTGCGAGCTGGGATCCCGACCCCGCGGCTGA 973 QY  
2387 ACAAGGCGACCCAGGACTTCTGGAGGTGCGAGCTGGGATCCCGACCCCGCGGCTGA 2446 Db  
974 AGAAGAGAGAGCGTGACCGTGTGAGCGTGGCGAGCGCTTACAGCGTGGCCCTGG 1033 QY  
2447 AGAAGAGAGAGCGTGACCGTGTGAGCGTGGCGAGCGCTTACAGCGTGGCCCTGG 2506 Db  
1034 ACAGGACTTCCGCAAGTACACCGCTTCAACCATCCCGAGCATCAACAAACGAGACCCCG 1093 QY  
2507 ACAGGACTTCCGCAAGTACACCGCTTCAACCATCCCGAGCATCAACAAACGAGACCCCG 2566 Db  
1094 GCATCGCTACAGTACAGCTGTGCTCCCGAGGCTGGAGGCGAGCGCCAGCATCTTCC 1153 QY  
2567 GCATCGCTACAGTACAGCTGTGCTCCCGAGGCTGGAGGCGAGCGCCAGCATCTTCC 2626 Db  
1154 AGAGCAGCATGACCAAGATCTTGGAGCCCTTCGCGCGCGCAACCCCGAGATCGTGATCT 1213 QY  
2627 AGAGCAGCATGACCAAGATCTTGGAGCCCTTCGCGCGCGCAACCCCGAGATCGTGATCT 2686 Db  
1214 ACCAGGCGCCCTGTGATCGTGGGAGCGAGCTGGAGATTCGGCGAGCAGCCCGCGAGATCG 1273 QY  
2687 ACCAGGCGCCCTGTGATCGTGGGAGCGAGCTGGAGATTCGGCGAGCAGCCCGCGAGATCG 2746 Db

1274 AGGAGCTGCGCAAGCACCTGTGCTGGGCTTCAACACCCCGCAACAGAGCACCGA 1333 QY  
2747 AGGAGCTGCGCAAGCACCTGTGCTGGGCTTCAACACCCCGCAACAGAGCACCGA 2806 Db  
1334 AGGAGCCCGCTTCTGTGTGATGGGTACGAGCTGCGACCCCGCAACAGTGGACCGTGCAGC 1393 QY  
2807 AGGAGCCCGCTTCTGTGTGATGGGTACGAGCTGCGACCCCGCAACAGTGGACCGTGCAGC 2860 Db  
1394 CCATCGAGCTGCGCGAGAGAGAGCTGACACCGTGAACGACATCCAGAGCTGTGGGCA 1453 QY  
2861 CCATCGAGCTGCGCGAGAGAGAGCTGACACCGTGAACGACATCCAGAGCTGTGGGCA 2920 Db  
1454 AGCTGAACTGGGCGAGCGAGATCTTACCCCGGATCAAGGTGGCGGAGCTGCGAGCTGC 1513 QY  
2921 AGCTGAACTGGGCGAGCGAGATCTTACCCCGGATCAAGGTGGCGGAGCTGCGAGCTGC 2980 Db  
1514 TCGCGCGCGCAAGGCCCTGACCCGACATCGTGTGCCCTGACCGAGGAGCCGAGCTGGAGC 1573 QY  
2981 TCGCGCGCGCAAGGCCCTGACCCGACATCGTGTGCCCTGACCGAGGAGCCGAGCTGGAGC 3040 Db  
1574 TGGCCGAGAACCGCGAGATCTTTCGCGAGCGCGTGCACGGGTGTACTACGACCCGAGCA 1633 QY  
3041 TGGCCGAGAACCGCGAGATCTTTCGCGAGCGCGTGCACGGGTGTACTACGACCCGAGCA 3100 Db  
1634 AGGACTGTGTGGCGGAGATCCAGAGAGGCGGCGACAGCAGTGGACCTTACAGATCTTACC 1693 QY  
3101 AGGACTGTGTGGCGGAGATCCAGAGAGGCGGCGACAGCAGTGGACCTTACAGATCTTACC 3160 Db  
1694 AGGAGCCCTTCAAGAACCTGAAAGACCGGCAAGTACGCGCAAGATGCGACCGGCCACCA 1753 QY  
3161 AGGAGCCCTTCAAGAACCTGAAAGACCGGCAAGTACGCGCAAGATGCGACCGGCCACCA 3220 Db  
1754 AGGACTGTGTGGCGGAGATCCAGAGAGGCGGCGACAGCAGTGGACCTTACAGATCTTACC 1813 QY  
3221 AGGACTGTGTGGCGGAGATCCAGAGAGGCGGCGACAGCAGTGGACCTTACAGATCTTACC 3280 Db  
1814 GGGGCAAGACCCCGAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGA 1873 QY  
3281 GGGGCAAGACCCCGAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGA 3340 Db  
1874 CGGACTGTGTGGCGGCGACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCGCTGG 1933 QY  
3341 CGGACTGTGTGGCGGCGACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCGCTGG 3400 Db  
1934 TGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGG 1993 QY  
3401 TGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGG 3460 Db  
1994 ACGGCGCGCGCAACCGCGAGACCAAGATCGGCAAGCCCGCTACGTGAACCGCGGCGC 2053 QY  
3461 ACGGCGCGCGCAACCGCGAGACCAAGATCGGCAAGCCCGCTACGTGAACCGCGGCGC 3520 Db  
2054 GGCAGAGATCGTGTGAGCTTGAACCGAGACCAACAGAGAGCGAGCTGCGAGGCGCATCC 2113 QY  
3521 GGCAGAGATCGTGTGAGCTTGAACCGAGACCAACAGAGAGCGAGCTGCGAGGCGCATCC 3580 Db  
2114 AGCTGGCGCTGCGAGGACAGCGCGAGAGGTGTACTGTGAGCTGGGTGCCCGCCAGAGGATCG 2173 QY  
3581 AGCTGGCGCTGCGAGGACAGCGCGAGAGGTGTACTGTGAGCTGGGTGCCCGCCAGAGGATCG 3640 Db  
2174 TGGGCGATCATCCAGGCGCGAGCCGCGAGAGCGAGCTGGTGAACAGATCATCG 2233 QY  
3641 TGGGCGATCATCCAGGCGCGAGCCGCGAGAGCGAGCTGGTGAACAGATCATCG 3700 Db  
2234 AGGAGCTGTGTGAGAGAGAGGTGTACTGTGAGCTGGGTGCCCGCCAGAGGATCG 2293 QY  
3701 AGGAGCTGTGTGAGAGAGAGGTGTACTGTGAGCTGGGTGCCCGCCAGAGGATCG 3760 Db  
2294 GCGGCAACGAGCAGATCGACAAAGCTGTGTGAGAGGAGGAGCTCCGCAAGGTGTGTCTTCTGG 2353 QY  
3761 GCGGCAACGAGCAGATCGACAAAGCTGTGTGAGAGGAGGAGCTCCGCAAGGTGTGTCTTCTGG 3820 Db  
2354 ACGGCATCGATGGCGGCGATCGTGTACTACAGTATACAGGAGCGAGCTTACGTGGGCGAGCG 2413 QY

Db 3821 ACGGCATCGATGGCGGCGATCGTGATCTACCAAGTACAGGACCTGTACGTGGCAGCG 3880  
 QY 2414 GCGGCGCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2457  
 Db 3881 GCGGCGCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 3924

RESULT 13  
 ACA03591  
 ID ACA03591 standard; DNA; 5184 BP.  
 AC ACA03591;  
 XX  
 XX  
 DT 22-MAY-2003 (first entry)  
 XX  
 DE Synthetic DNA encoding immunogenic HIV peptide #74.  
 XX  
 KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
 KW gene therapy; packaging cell line; humoral immune response;  
 KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
 XX  
 OS Synthetic.  
 XX  
 XX WO2003004657-A1.  
 XX  
 XX 16-JAN-2003.  
 XX  
 XX 05-JUL-2002; 2002WO-US021421.  
 XX  
 XX 05-JUL-2001; 2001US-0303192P.  
 XX  
 XX 31-AUG-2001; 2001US-0316860P.  
 PR  
 PR 16-JAN-2002; 2002US-0349728P.  
 PR  
 PR 16-JAN-2002; 2002US-0349793P.  
 PR  
 PR 16-JAN-2002; 2002US-0349871P.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 XX  
 XX Zur Megede J, Barnett SW, Lian Y;  
 XX  
 XX WPI; 2003-221602/21.  
 XX  
 XX  
 PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
 PT polypeptides, useful as immunogenic compositions or vaccines for  
 PT generating humoral or cellular immune responses against HIV in a subject,  
 PT especially humans.  
 XX  
 XX Example 1; Fig 79; 262pp; English.  
 XX  
 XX The invention describes a synthetic polynucleotide encoding 2 or more  
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
 CC derived from different HIV subtypes. The polynucleotide is useful for  
 CC immunisation, generation of packaging cell lines, or production of HIV  
 CC polypeptides. The polynucleotide and its encoded proteins are useful as  
 CC immunogenic compositions or vaccines for generating humoral or cellular  
 CC immune responses against HIV in a subject, or for inducing neutralising  
 CC antibodies against HIV. The gene delivery vector comprising the  
 CC polynucleotide is also useful for DNA immunisation of, or for generating  
 CC an immune response (e.g. a humoral or cellular immune response), in, a  
 CC subject such as a mammal, particularly a human. This sequence encodes a  
 CC human immunodeficiency virus immunogenic peptide  
 XX  
 XX Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;  
 XX  
 XX Query Match 98.0%; Score 2414; DB 7; Length 5184;  
 XX Best Local Similarity 99.5%; Pred. No. 3e-291;  
 XX Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;  
 QY 14 TGCGGAGCCATGAGCCAGCCACCGCCCAATCTCTGATGCGGAGCAACTTCA 73  
 Db 2741 TGCGGAGCCATGAGCCAGCCACCGCCCAATCTCTGATGCGGAGCAACTTCA 2800  
 QY 74 AGGGCCCCAAGCGCATCATCAAGTCTTCACTCGCGCAAGGAGGGCCCATCGCCGCA 133

2801 AGGCCCCCAAGCGCATCATCAAGTCTTCAATGCGCGAAGGAGGGCCCAATCGGCCGCA 2860  
 QY 134 ACTCCCGCGCCCGCCCAAGAGGGTGTGGAAGTGGCGCAAGGAGGCCACCAAGATGA 193  
 Db 2861 ACTGCGCGCCCGCCCAAGAGGGTGTGGAAGTGGCGCAAGGAGGGCCACCAAGATGA 2920  
 QY 194 AGGACTGCAACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCTTCCCGCAGGCA 253  
 Db 2921 AGGACTGCAACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCTTCCCGCAGGCA 2980  
 QY 254 AGGCGCGCGAGTTCCCGCAGGAGCAGAAACCGCGCCCAACAGCCCCCAGCCGAGCTGC 313  
 Db 2981 AGGCGCGCGAGTTCCCGCAGGAGCAGAAACCGCGCCCAACAGCCCCCAGCCGAGCTGC 3040  
 QY 314 AGGTGCGCGCGAGCAACCGCCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
 Db 3041 AGGTGCGCGCGAGCAACCGCCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3100  
 QY 374 TCCCGCAGATCACCGCTGTGCGAGCGCCCGCTGTGAGCATCAAGGTGGCGCGCGAGATCA 433  
 Db 3101 TCCCGCAGATCACCGCTGTGCGAGCGCGCGCTGTGAGCATCAAGGTGGCGCGCGAGATCA 3160  
 QY 434 AGGAGCGCGCTGTGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493  
 Db 3161 AGGAGCGCGCTGTGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3220  
 QY 494 GCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAAGGTGCGCGCGAGTACG 553  
 Db 3221 GCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAAGGTGCGCGCGAGTACG 3280  
 QY 554 ACAGATCTCTGATCGAGATCTGCGCGCAAGAGCCCATCGCGCGCGCGCGCGCGCGCGCG 613  
 Db 3281 ACCAGATCTCTGATCGAGATCTGCGCGCAAGAGCCCATCGCGCGCGCGCGCGCGCGCGCG 3340  
 QY 614 CCCCCGTGAACATCATCGCGCGCAACATGTCACCGAGCTGGCGCTGCAACCTTGAACCTTCC 673  
 Db 3341 CCCCCGTGAACATCATCGCGCGCAACATGTCACCGAGCTGGCGCTGCAACCTTGAACCTTCC 3400  
 QY 674 CCATCAGCCCCATCGAGAGCGCGTCCGTAAGCTGAAGCCCGCGCATGAGCGGCCCAAGG 733  
 Db 3401 CCATCAGCCCCATCGAGAGCGCGTCCGTAAGCTGAAGCCCGCGCATGAGCGGCCCAAGG 3460  
 QY 734 TGAAGCAGTGGCGCGCTGACCGCGAGAGATCAAGGCCCTGACCGCGCGCGCGCGCGCGCGCG 793  
 Db 3461 TGAAGCAGTGGCGCGCTGACCGCGAGAGATCAAGGCCCTGACCGCGCGCGCGCGCGCGCGCG 3520  
 QY 794 TGGAGAGGAGGCGCAAGATCAACCAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 853  
 Db 3521 TGGAGAGGAGGCGCAAGATCAACCAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3580  
 QY 854 TCGCCATCAAGAAAGAGGAGCAGCAACCAAGTGGCGCAAGCTGTGGACTTTCGCGAGCTGA 913  
 Db 3581 TCGCCATCAAGAAAGAGGAGCAGCACCAGTGGCGCAAGCTGTGGACTTTCGCGAGCTGA 3640  
 QY 914 ACAGCGCACCCAGGACTTCTGGAGGTGCGAGTGGCGCATCCCCCAGCGCGCGCGCGCGCGCG 973  
 Db 3641 ACAGCGCACCCAGGACTTCTGGAGGTGCGAGTGGCGCATCCCCCAGCGCGCGCGCGCGCGCG 3700  
 QY 974 AGAAGAAAGAGCGTGAACCGTCTGAGCGTGGCGAGCGCGCTACTTTCAGCGTGGCGCGCGCG 1033  
 Db 3701 AGAAGAAAGAGCGTGAACCGTCTGAGCGTGGCGAGCGCGCTACTTTCAGCGTGGCGCGCGCG 3760  
 QY 1034 ACAGGAGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAGAGAGAGAGAGAGAG 1093  
 Db 3761 ACAGGAGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAGAGAGAGAGAGAGAG 3820  
 QY 1094 GCATCCGCTACCAAGTACACCGTCTGCCCGAGGGCTGGAAGGGCGAGCGCGCGCGCGCGCGCG 1153  
 Db 3821 GCATCCGCTACCAAGTACACCGTCTGCCCGAGGGCTGGAAGGGCGAGCGCGCGCGCGCGCGCG 3880  
 QY 1154 AGAGCAGATGACCAAGATCCTGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1213



Db 2741 TCGCCGAGGCCATGAGCCAGGCCACCGCGCCAAACATCTCTGATGCGAGCGCAACTTCA 2800  
 Qy 74 AGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGGGCAAGAGGGCCACATCGCCCGCA 133  
 Db 2801 AGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGGGCAAGAGGGCCACATCGCCCGCA 2860  
 Qy 134 ACTCGCGGCCCCCGCGCAAGAGGGCTGTGAAAGTGGGCAAGAGGGCCACCAAGATGA 193  
 Db 2861 ACTCGCGGCCCCCGCGCAAGAGGGCTGTGAAAGTGGGCAAGAGGGCCACCAAGATGA 2920  
 Qy 194 AGGACTGCACGAGCGCGCCAGGCAACTTCTCCGCGAGGACCTGGCCCTTCCGCCAGGGCA 253  
 Db 2921 AGGACTGCACGAGCGCGCCAGGCAACTTCTCCGCGAGGACCTGGCCCTTCCGCCAGGGCA 2980  
 Qy 254 AGGCGCGAGTTCCTCCAGCGAGCAACCGCGCCAAACAGCCCCCAACGAGCGCGAGCTGC 313  
 Db 2981 AGGCGCGAGTTCCTCCAGCGAGCAACCGCGCCAAACAGCCCCCAACGAGCGCGAGCTGC 3040  
 Qy 314 AGGTGCGCGGACCAACCCCGCGAGCGCGCGCGCGAGCGCGAGCGCGAGCTTGAACCT 373  
 Db 3041 AGGTGCGCGGACCAACCCCGCGAGCGCGCGCGCGAGCGCGAGCGCGAGCTTGAACCT 3100  
 Qy 374 TCCCCCAGATACCTCTGTGGCAGCGCCCTCTGTGAGCATCAAGGTGGCGCGCGAGATCA 433  
 Db 3101 TCCCCCAGATACCTCTGTGGCAGCGCCCTCTGTGAGCATCAAGGTGGCGCGCGAGATCA 3160  
 Qy 434 AGGAGCCCTGCTGACACCGCGCGCGAGCACACCGTGTGAGGAGATGAGCCCTGCCCG 493  
 Db 3161 AGGAGCCCTGCTGACACCGCGCGCGAGCACACCGTGTGAGGAGATGAGCCCTGCCCG 3220  
 Qy 494 GCAAGTGGAGCCCAAGATGATCGCGGCGATCGCGGGCTTCATCAAGGTGGCGCGAGTACG 553  
 Db 3221 GCAAGTGGAGCCCAAGATGATCGCGGCGATCGCGGGCTTCATCAAGGTGGCGCGAGTACG 3280  
 Qy 554 ACCAGATCTGATCGAGATCTGGGCAAGAGGCCATCGGCACCGTGTGATCGCGGCCCA 613  
 Db 3281 ACCAGATCTGATCGAGATCTGGGCAAGAGGCCATCGGCACCGTGTGATCGCGGCCCA 3340  
 Qy 614 CCCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTCC 673  
 Db 3341 CCCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTCC 3400  
 Qy 674 CCATAGCCCCCATCGAGACCGTCCGCTGAGCTGAGCTGAGCCCGGCGATGAGCGGCCCAAG 733  
 Db 3401 CCATAGCCCCCATCGAGACCGTCCGCTGAGCTGAGCTGAGCCCGGCGATGAGCGGCCCAAG 3460  
 Qy 734 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGGCATCTCGGAGGAGA 793  
 Db 3461 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGGCATCTCGGAGGAGA 3520  
 Qy 794 TGAAGAGGAGGCAAGATCAACAGATCGGCCCCCGAGAACCCCTTACAAACCCCCCGTGT 853  
 Db 3521 TGAAGAGGAGGCAAGATCAACAGATCGGCCCCCGAGAACCCCTTACAAACCCCCCGTGT 3580  
 Qy 854 TCGCCATCAAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGA 913  
 Db 3581 TCGCCATCAAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGA 3640  
 Qy 914 ACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGATTCGCCCAACCCCGCGCGCTGA 973  
 Db 3641 ACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGATTCGCCCAACCCCGCGCGCTGA 3700  
 Qy 974 AGAGAGAGAGCGGTGACCGTGTGAGCGTGGCGAGCGCTTACGCGTGCCTCGG 1033  
 Db 3701 AGAGAGAGAGCGGTGACCGTGTGAGCGTGGCGAGCGCTTACGCGTGCCTCGG 3760  
 Qy 1034 ACAGGAGACTTCCGCAAGTACACCGCCCTTCCATCCCGAGCATCAACACAGAGACCCCG 1093  
 Db 3761 ACAGGAGACTTCCGCAAGTACACCGCCCTTCCATCCCGAGCATCAACACAGAGACCCCG 3820  
 Qy 1094 GCATCCGCTACAGTACAAAGTGTGCTGCCCGAGGCTGGAGGCGAGCCCGAGCTTCC 1153  
 Db 3821 GCATCCGCTACAGTACAAAGTGTGCTGCCCGAGGCTGGAGGCGAGCCCGAGCTTCC 3880

Qy 1154 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCGCGCAACCCCGAGATCTGATCT 1213  
 Db 3881 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCGCGCAACCCCGAGATCTGATCT 3940  
 Qy 1214 ACCAGAGCCCTCTGACGTGGGCGAGCGACTCGAGATCGGCGAGCACCGCGCAAGATCG 1273  
 Db 3941 ACCAGAGCCCTCTGACGTGGGCGAGCGACTCGAGATCGGCGAGCACCGCGCAAGATCG 4000  
 Qy 1274 AGAGTGTGCGCAAGCACTGTCTGGGCTTCAACACCCCGCAAGAGACCAAG 1333  
 Db 4001 AGAGTGTGCGCAAGCACTGTCTGGGCTTCAACACCCCGCAAGAGACCAAG 4060  
 Qy 1334 AGAGAGCCCTCTCTGTGGATGGGCTGAGCTGACCGCGCGCAAGTGGAGCTGCGAGC 1393  
 Db 4061 AGAGAGCCCTCTCTGTGGATGGGCTGAGCTGACCGCGCGCAAGTGGAGCTGCGAGC 4114  
 Qy 1394 CCATCGAGCTGCGCGAGAGAGAGTGTGACCGTGAACGACATCCAGAAAGCTGTGGGCA 1453  
 Db 4115 CCATCGAGCTGCGCGAGAGAGAGTGTGACCGTGAACGACATCCAGAAAGCTGTGGGCA 4174  
 Qy 1454 AGTGAACCTGGGCGAGCGAGTGTACCCCGGATCAAGGTGGCGCGAGCTGTGCAAGCTGC 1513  
 Db 4175 AGTGAACCTGGGCGAGCGAGTGTACCCCGGATCAAGGTGGCGCGAGCTGTGCAAGCTGC 4234  
 Qy 1514 TGGCGCGCGCAAGAGCCCTGACCGACATCTGTGCCCTGACCGAGGAGGCGCGAGCTGAGC 1573  
 Db 4235 TGGCGCGCGCAAGAGCCCTGACCGACATCTGTGCCCTGACCGAGGAGGCGCGAGCTGAGC 4294  
 Qy 1574 TGGCGGAGAACCGCGAGATCTCTCGGAGCGCGTGCACGGCGTGTACTAGACCCCGAGCA 1633  
 Db 4295 TGGCGGAGAACCGCGAGATCTCTCGGAGCGCGTGCACGGCGTGTACTAGACCCCGAGCA 4354  
 Qy 1634 AGGACCTGTGGCGCGAGATCCAGAAAGCGGCGCACGACAGTGGAGCTTACAGATCTAC 1693  
 Db 4355 AGGACCTGTGGCGCGAGATCCAGAAAGCGGCGCACGACAGTGGAGCTTACAGATCTAC 4414  
 Qy 1694 AGGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACCGCAAGTGGCGAGCGCGCCACCA 1753  
 Db 4415 AGGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACCGCAAGTGGCGAGCGCGCCACCA 4474  
 Qy 1754 ACAGCTGTGAAGCAGCTGACCGAGCGCGTGCAGAAAGTTCGCGTGGAGAGCATCTGATCT 1813  
 Db 4475 ACAGCTGTGAAGCAGCTGACCGAGCGCGTGCAGAAAGTTCGCGTGGAGAGCATCTGATCT 4534  
 Qy 1814 GGGCAAGACCCCGAAGTTCCGCTGCGCATCCAGAAAGAGACCTGGAGACCTGGTGA 1873  
 Db 4535 GGGCAAGACCCCGAAGTTCCGCTGCGCATCCAGAAAGAGACCTGGAGACCTGGTGA 4594  
 Qy 1874 CCGACTACTGGCGGCGCACCTGATCCCGAGTGGGAGTTCTGTGAACACCCCGCGCTGG 1933  
 Db 4595 CCGACTACTGGCGGCGCACCTGATCCCGAGTGGGAGTTCTGTGAACACCCCGCGCTGG 4654  
 Qy 1934 TGAAGCTGTGGTACCAAGTGGAGAGGCGCATCATTCGCGCGCGAGACCTTCTACGTGG 1993  
 Db 4655 TGAAGCTGTGGTACCAAGTGGAGAGGCGCATCATTCGCGCGCGAGACCTTCTACGTGG 4714  
 Qy 1994 ACAGCGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAGCGCGGCC 2053  
 Db 4715 ACAGCGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAGCGCGGCC 4774  
 Qy 2054 GGCAGAAAGATCTGTGAGCTTGAACCGAGACCAACCAAGAGACCGAGCTGCGAGCCATCC 2113  
 Db 4775 GGCAGAAAGATCTGTGAGCTTGAACCGAGACCAACCAAGAGACCGAGCTGCGAGCCATCC 4834  
 Qy 2114 AGCTGGCCCTGCGAGGACAGCGGCGAGGAGTGAACATCTGTGACCGAGCGAGCGAGTACGCC 2173  
 Db 4835 AGCTGGCCCTGCGAGGACAGCGGCGAGGAGTGAACATCTGTGACCGAGCGAGCGAGTACGCC 4894  
 Qy 2174 TGGGCAATCATCCAGGCCCGAGCAAGAGCGAGAGCGAGCTGGTGAACACAGATCATCTCG 2233  
 Db 4895 TGGGCAATCATCCAGGCCCGAGCAAGAGCGAGAGCGAGCTGGTGAACACAGATCATCTCG 4954

QY 2234 AGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCGCGCCCAAGGGGATCG 2293  
Db 4955 AGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCGCGCCCAAGGGGATCG 5014  
QY 2294 GCGGCAACGAGCAGATCGCAAGCTGTGTAGCAAGGAGCATCCGCAAGGTGCTTCTCTGG 2353  
Db 5015 GCGGCAACGAGCAGATCGCAAGCTGTGTAGCAAGGAGCATCCGCAAGGTGCTTCTCTGG 5074  
QY 2354 ACGGCATCGATGGCGGATCGTGATCTACCACTACATGAGACGACCTGTACGTGGGAGCG 2413  
Db 5075 ACGGCATCGATGGCGGATCGTGATCTACCACTACATGAGACGACCTGTACGTGGGAGCG 5134  
QY 2414 GCGGCCCTAGCATGATTAAGGCTTCCGCGGCTAGCACCGGT 2457  
Db 5135 GCGGCCCTAGCATGATTAAGGCTTCCGCGGCTAGCACCGGT 5178

RESULT 15  
ID ADCL13234  
XX ADCL13234 standard; DNA; 3531 BP.  
AC ADCL13234;  
XX  
XX  
DT 18-DEC-2003 (first entry)  
XX  
XX DNA of HIV construct GagPolmut\_C SEQ ID NO 13.  
XX  
XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
XX Rev; HIV type C; Gene therapy; vaccine; immunisation; HIV; ds.  
XX Human immunodeficiency virus.  
XX OS  
XX WO2003004620-A2.  
XX  
XX 16-JAN-2003.  
XX  
XX 05-JUL-2002; 2002WO-US021420.  
XX  
XX 05-JUL-2001; 2001US-0303192P.  
XX 31-AUG-2001; 2001US-0316860P.  
XX 16-JAN-2002; 2002US-0349871P.  
XX  
XX (CHIR ) CHIRON CORP.  
XX (UYST-) UNIV STELLENBOSCH.  
XX  
XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
XX WPI; 2003-221593/21.  
XX  
XX New expression cassette comprising a polynucleotide sequence encoding a  
XX polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
XX Prot, or Rev polypeptide, useful for immunization, or generating  
XX packaging cell lines.  
XX  
XX Disclosure; Fig 10; 301pp; English.  
XX  
XX The invention relates to a novel expression cassette comprising a  
XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
XX Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
XX expression cassette can be used to treat HIV type C by gene therapy or  
XX used in the development of a vaccine. The gene delivery vector is  
XX administered intramuscularly, intramusosally, intranasally,  
XX subcutaneously, intradermally, transdermally, intravaginally,  
XX intrarectally, orally or intravenously. The expression cassette is useful  
XX for immunisation, generating packaging cell lines and producing HIV  
XX polypeptides. This polynucleotide sequence represents the DNA of an HIV  
XX Type C related sequence of the invention.  
XX  
XX Sequence 3531 BP; 802 A; 1210 C; 1096 G; 423 T; 0 U; 0 Other;  
SQ

Query Match 96.8%; Score 2383.6; DB 9; Length 3531;  
Best Local Similarity 98.8%; Pred. No. 1.9e-287;  
Matches 2414; Conservative 0; Mismatches 24; Indels 6; Gaps 1;

QY 14 TGGCCGAGGCGCATGAGCGAGCCACAGCGCCCAACATCCTGTATGCGAGCGCAACTTCA 73  
Db 1088 TGGCCGAGGCGCATGAGCGAGCCACAGCGCCCAACATCCTGTATGCGAGCAACTTTTAAA 1147  
QY 74 AGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTCGCGCAAGAGGGGCCACATCGCCCGCA 133  
Db 1148 AGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTCGCGCAAGAGGGGCCACATCGCCCGCA 1207  
QY 134 ACTGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGGAGAGGGGCCACAGATGA 193  
Db 1208 ACTGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGGAGAGGGGCCACAGATGA 1267  
QY 194 AGGACTGCACCGAGCGCGCAGGCCCAACTTCTTCGCCGAGAGCACTGGCCTTCCCCCAGGGCA 253  
Db 1268 AGGACTGCACCGAGCGCGCAGGCCCAACTTCTTCGCCGAGAGCACTGGCCTTCCCCCAGGGCA 1327  
QY 254 AGGCCCGCGAGTTCCCGAGCGAGCAGAAACCGCGCCCAACAGCCCCCAACAGCGCGAGTGTC 313  
Db 1328 AGGCCCGCGAGTTCCCGAGCGAGCAGAAACCGCGCCCAACAGCCCCCAACAGCGCGAGTGTC 1387  
QY 314 AGGTGCGCGGCGACAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
Db 1388 AGGTGCGCGGCGACAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1447  
QY 374 TCCCCCAGATCACCTGTGTGCGAGCGCCCCCTGTGTGAGCATCAAGTGTGGCGGCGCAGATCA 433  
Db 1448 TCCCCCAGATCACCTGTGTGCGAGCGCCCCCTGTGTGAGCATCAAGTGTGGCGGCGCAGATCA 1507  
QY 434 AGGAGCCCTGTGAGCACCG 493  
Db 1508 AGGAGCCCTGTGTGAGCACCG 1567  
QY 494 GCAAGTGAAGCCCAAGATGATCGCGCGCGCATCGCGCGGCTTTCATCAAGTGTGGCGCGCAGTACG 553  
Db 1568 GCAAGTGAAGCCCAAGATGATCGCGCGCGCATCGCGCGGCTTTCATCAAGTGTGGCGCGCAGTACG 1627  
QY 554 ACCAGATCCTGTATGAGATCTTGGCGCAAGAGAGGCATCGGCACCGTGTGTATCGGCCCGCA 613  
Db 1628 ACCAGATCCTGTATGAGATCTTGGCGCAAGAGAGGCATCGGCACCGTGTGTATCGGCCCGCA 1687  
QY 614 CCCCCGTGAACATCATCGCGCGCGCAACATCTGACCCAGCTGGGTGCGCACCCCTGAACTTCC 673  
Db 1688 CCCCCGTGAACATCATCGCGCGCGCAACATCTGACCCAGCTGGGTGCGCACCCCTGAACTTCC 1747  
QY 674 CCATCAGCCCCATCGAGACCGTGCCTGTAAGCTGTAAGCCCGGATGAGACGGCCCGCAAGG 733  
Db 1748 CCATCAGCCCCATCGAGACCGTGCCTGTAAGCTGTAAGCCCGGATGAGACGGCCCGCAAGG 1807  
QY 734 TGAACAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTGCGAGGAGA 793  
Db 1808 TGAACAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTGCGAGGAGA 1867  
QY 794 TGGAGAAGGAGGGCAAGATCAACAGATCGGCGCGCGAGAACCCCTTACAAACCCCCCGTGT 853  
Db 1868 TGGAGAAGGAGGGCAAGATCAACAGATCGGCGCGCGAGAACCCCTTACAAACCCCCCGTGT 1927  
QY 854 TCGCCATCAAGAGAGAGACAGCACCAGTGGCGCGAGCTGTGGACTTCGCGGAGCTGA 913  
Db 1928 TCGCCATCAAGAGAGAGACAGCACCAGTGGCGCGAGCTGTGGACTTCGCGGAGCTGA 1987  
QY 914 ACAACGCAACCCAGGACTTCTTGGGAGGTGAGCTGGGCATCCCCCAGCCCGCGGCTTGA 973  
Db 1988 ACAACGCAACCCAGGACTTCTTGGGAGGTGAGCTGGGCATCCCCCAGCCCGCGGCTTGA 2047  
QY 974 AGAAGAAGAGAGCTGACCGTGTGTGAGCGTGGCGGAGCGCTACTTACAGCGTGCCTTGG 1033  
Db 2048 AGAAGAAGAGAGCTGACCGTGTGTGAGCGTGGCGGAGCGCTACTTACAGCGTGCCTTGG 2107  
QY 1034 ACGAGGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAACGAGACCCCCCG 1093  
Db 2108 ACGAGGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAACGAGACCCCCCG 2167

QY	1094	GCATCGCTACAGTACAACTGCTGCCCCAGGGCTGGAAGGCGAGCCCGACATCTTCC	1153
Db	2168	GCATCGCTACAGTACAACTGCTGCCCCAGGGCTGGAAGGCGAGCCCGACATCTTCC	2227
QY	1154	AGAGCGATGACCAAGATCTCGAGCCCTTCGGCGCCGCGAACCOCAGATCTGTATCT	1213
Db	2228	AGAGCGATGACCAAGATCTCGAGCCCTTCGGCGCCGCGAACCOCAGATCTGTATCT	2287
QY	1214	ACCAGGCCCTCTGTACGTGGGCGAGCGACTCGAGATCGGCCAGCACCGCGCCAGATCG	1273
Db	2288	ACCAGGCCCTCTGTACGTGGGCGAGCGACTCGAGATCGGCCAGCACCGCGCCAGATCG	2347
QY	1274	AGAGCTGGCGAAGCACTCTGCGTGGGCTTCAACCCCGCAAGAGCAACCA	1333
Db	2348	AGAGCTGGCGAAGCACTCTGCGTGGGCTTCAACCCCGCAAGAGCAACCA	2407
QY	1334	AGAGGCCCTCTGTGTGGTACGGCTACGGCTGACCCCGCACCAAGTGGACCGTGCAGC	1393
Db	2408	AGAGGCCCTCTCTCTGCCAT-----CGAGCTGACCCCGCAAGTGGACCGTGCAGC	2461
QY	1394	CCATCGAGCTGCCCGAAGAGAGAGTGGACCGTGAACGACATCCAGAGCTGGTGGCA	1453
Db	2462	CCATCGAGCTGCCCGAAGAGAGAGTGGACCGTGAACGACATCCAGAGCTGGTGGCA	2521
QY	1454	AGCTGAATGGGCGAGCGAGTACTCCCGGATCAAGTGGCGCAGCTGTCAAGCTGC	1513
Db	2522	AGCTGAATGGGCGAGCGAGTACTCCCGGATCAAGTGGCGCAGCTGTCAAGCTGC	2581
QY	1514	TGCGCGCGCCAAAGGCCCTGACCGACATCGTGCCTCTGACCGAGGCGCGAGCTGGAC	1573
Db	2582	TGCGCGCGCCAAAGGCCCTGACCGACATCGTGCCTCTGACCGAGGCGCGAGCTGGAC	2641
QY	1574	TGCGCGAGACCGCGAGTCTCGCGAGCCCGTGCACCGCGTGTACTACGACCCCGCA	1633
Db	2642	TGCGCGAGACCGCGAGTCTCGCGAGCCCGTGCACCGCGTGTACTACGACCCCGCA	2701
QY	1634	AGACCTGTGTGGCGAGATCCAGAGCAGGCGCCACGACAGTGGACCTACAGATCTACC	1693
Db	2702	AGACCTGTGTGGCGAGATCCAGAGCAGGCGCCACGACAGTGGACCTACAGATCTACC	2761
QY	1694	AGAGCCCTTCAAGACCTGAAGACCGGCAAGTACGCGAGATGCGACCGGCCACCA	1753
Db	2762	AGAGCCCTTCAAGACCTGAAGACCGGCAAGTACGCGAGATGCGACCGGCCACCA	2821
QY	1754	ACGACGTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAGAGCATCTGTATCT	1813
Db	2822	ACGACGTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAGAGCATCTGTATCT	2881
QY	1814	GGGCAAGACCCCGAGTTCGCGCTGCCATCCAGAGAGAGCTGGGAGACCTGTGGA	1873
Db	2882	GGGCAAGACCCCGAGTTCGCGCTGCCATCCAGAGAGAGCTGGGAGACCTGTGGA	2941
QY	1874	CCGACTACTGGCAGGCCACTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGG	1933
Db	2942	CCGACTACTGGCAGGCCACTGTGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGG	3001
QY	1934	TGAAGCTGTGTACCACTGGAGAGAGGCCATCATTCGGCGCGAGAGCTTCTACGTGG	1993
Db	3002	TGAAGCTGTGTACCACTGGAGAGAGGCCATCATTCGGCGCGAGAGCTTCTACGTGG	3061
QY	1994	ACGGCGCGCCAAACCGGAGACCAAGATCGCGAGCGCGCTAGCTGACCGCGGGCC	2053
Db	3062	ACGGCGCGCCAAACCGGAGACCAAGATCGCGAGCGCGCTAGCTGACCGCGGGCC	3121
QY	2054	GGCAGAAGATCGTGAAGCTGACCGAGACCAACCAAGAGAGCGAGCTGACGGCCATCC	2113
Db	3122	GGCAGAAGATCGTGAAGCTGACCGAGACCAACCAAGAGAGCGAGCTGACGGCCATCC	3181
QY	2114	AGCTGGCCCTGCAGGACAGCGGAGGAGTGTACATCGTGAACGAGCGAGCTAGCGCC	2173
Db	3182	AGCTGGCCCTGCAGGACAGCGGAGGAGTGTACATCGTGAACGAGCGAGCTAGCGCC	3241
QY	2174	TGGGCATCATCCAGGCCCGCCGACCAAGAGCGAGAGCTGGTGAACAGATCATCG	2233

Search completed: April 10, 2004, 07:32:28  
Job time : 629 secs

Db	3242	TGGGCATCATCCAGGCCCGCCGACAGAGCGAGAGCTGGTGAACAGATCATCG	3301
QY	2234	AGCAGCTGATCAAGAAGGAGAAAGTGTACTTGAAGTGGGTGCCCGCCACAAAGGCGATCG	2293
Db	3302	AGCAGCTGATCAAGAAGGAGAAAGTGTACTTGAAGTGGGTGCCCGCCACAAAGGCGATCG	3361
QY	2294	GCGGCAACGAGCGAGATCGAACAGCTGGTGAAGAGGGCATCCGCAAGGTGCTTCTCTGG	2353
Db	3362	GCGGCAACGAGCGAGATCGAACAGCTGGTGAAGAGGGCATCCGCAAGGTGCTTCTCTGG	3421
QY	2354	ACGGCATCGATGGCGGCATCGTATCTACAGTACATGACGACCTGTACGTGGGCGAGCG	2413
Db	3422	ACGGCATCGATGGCGGCATCGTATCTACAGTACATGACGACCTGTACGTGGGCGAGCG	3481
QY	2414	GCGGCCCTAGGATCGAATTAAGCTTCCCGGGGCTAGCACCCGT	2457
Db	3482	GCGGCCCTAGGATCGAATTAAGCTTCCCGGGGCTAGCACCCGT	3525



Sequence 5, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 15, Appl  
Sequence 15, Appl  
Sequence 15, Appl  
Sequence 13, Appl  
Sequence 13, Appl  
Sequence 33, Appl  
Sequence 33, Appl  
Sequence 35, Appl  
Sequence 35, Appl  
Sequence 35, Appl  
Sequence 74, Appl  
Sequence 1, Appl  
Sequence 3, Appl  
Sequence 7, Appl  
Sequence 7, Appl  
Sequence 6, Appl  
Sequence 83, Appl  
Sequence 101, Appl

## ALIGNMENT'S

RESULT 1  
US-09-475-515-82  
: sequence 82: Application US/09475515A

GENERAL INFORMATION:

APPLICANT: BARNETT, Susan  
APPLICANT: ZUR MEGEDE, Jan  
APPLICANT: SRIVASTAVA, Indresh  
APPLICANT: LIAN, Ying  
APPLICANT: HARTOG, Karin  
APPLICANT: LIU, Hong  
APPLICANT: GREER, Catherine  
APPLICANT: SELBY, Mark

```

1  APPLICANT: WALKER, CHRISOPHER
2  TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
3  TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
4  FILE REFERENCE: 1621.002
5  CURRENT APPLICATION NUMBER: US/09/475.515A
6  CURRENT FILING DATE: 1999-12-30
7  NUMBER OF SEQ ID NOS: 90
8  SOFTWARE: Patent In Ver. 2.0
9

```

```

/ NAME:
/
/ LENGTH: 2306
/
/ TYPE: DNA
/
/ ORGANISM: Artificial Sequence
/
/ FEATURE:
/
/ OTHER INFORMATION: Description of Artificial Sequence:
/ OTHER INFORMATION: FS(-).protmod.RTopt.YM
US-09-475-515-82

```

Query Match 83.1%; Score 2046; DB 4; Length 2306;  
Best Local Similarity 93.6%; Pred. No. 0;

170	QY	CGCGCAAGAGGGGCCACCGATGAGAGGACTGCACCGAGCGCCAGGCCAACTCTCTCCGGC	229
		CGCGCGCGAAGGACACCAATGAAAGATTGCCTGAGAGCAGGCTAATTTCTTCCGGC	60
1	DB		
230	QY	AGGACCTTGGCTTCCCCCGGCGCAAGGCCGGGAGTTCCCGCAGCGAGCAGAACCCGCGCA	289
61	DB	AGGACCTTGGCTTCTCTCGAGGCAGGCCCGCGAGTTCAGCGCGGAGCAGACCCCGGCCA	120
290	QY	ACAGCCCCCACCAGCCGGCGAGCTCAGGTGCGGGCGG-----ACAACCCCGCCGACGAGG	343
121	DB	ACAGCCCCCACCAGCTGAGGTGCGGGCGGCGAGCTGAGGTGCGGGCGGCGAGCAACACCTTGAGCGGAG	180
344	QY	CGGGCGCGGAGCGCCAGGCAACCTTG-----AATCTCCCGCAGATCAACCTTGTCGAGC	397
181	DB	CGGGCGCGGAGCGCCAGGCAACCGTGAGCTTCAACTTCCCGCAGATCAACCTTGTCGAGC	240

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 06:48:32 ; Search time 118.333 Seconds  
(without alignments)  
11550.795 Million cell updates/sec

Title: US-09-610-313-31  
Perfect score: 2463  
Sequence: 1 atcagccgccaccatgcccga.....ggcctagcaccggtgaattc 2463

Scoring table: IDENTITY\_NUC  
Gapop 10.0 Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

[illegible]

**Post-processing:** Minimum Match 0%  
Maximum Match 100%

гастрування є грати

```
Database :
Issued_Attribute :
1: /cgn2_6/ptodata/2/ina/5A COMB seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB seq.*
5: /cgn2_6/ptodata/2/ina/PCRUS COMB seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	2046	83.1	2306	4	US-09-475-515-82	Sequence 82, Appl
2	2025.2	82.2	2312	4	US-09-475-515-84	Sequence 84, Appl
3	2019.2	82.0	2300	4	US-09-475-515-83	Sequence 83, Appl
4	1942.6	78.3	4319	4	US-09-475-515-6	Sequence 6, Appl
5	1878.8	76.9	2305	4	US-09-475-515-80	Sequence 80, Appl
6	1852	75.2	2299	4	US-09-475-515-81	Sequence 81, Appl
7	1651.8	67.1	4307	4	US-09-552-950-2	Sequence 2, Appl
8	1624.6	66.0	9772	4	US-09-552-950-5	Sequence 5, Appl
9	1556	63.6	8366	4	US-09-872-733A-6	Sequence 6, Appl
10	1530.6	62.1	4338	4	US-09-872-733A-1	Sequence 1, Appl
11	1203.6	48.9	9010	4	US-09-184-418C-8	Sequence 8, Appl
12	1172.6	47.6	8972	4	US-09-184-418C-9	Sequence 9, Appl
13	1165.4	47.3	8599	4	US-09-184-418C-11	Sequence 11, Appl
14	1163.8	47.3	2467	4	US-09-872-733A-3	Sequence 3, Appl
15	1142	46.4	8992	4	US-09-184-418C-4	Sequence 4, Appl
16	1106.8	44.9	2601	3	US-09-117-217-9	Sequence 7, Appl
17	1106.8	44.9	2601	3	US-09-117-217-9	Sequence 9, Appl
18	1106.8	44.3	2601	3	US-09-117-217-11	Sequence 11, Appl
19	1106.8	44.9	2601	3	US-09-117-217-13	Sequence 13, Appl
20	1106.8	44.9	2601	4	US-09-735-487-7	Sequence 7, Appl
21	1106.8	44.9	2601	4	US-09-735-487-9	Sequence 9, Appl
22	1106.8	44.9	2601	4	US-09-735-487-11	Sequence 11, Appl
23	1106.8	44.9	2601	4	US-09-735-487-13	Sequence 13, Appl
24	1106.8	44.9	4307	4	US-09-552-950-1	Sequence 1, Appl
25	1106.8	44.9	9719	4	US-09-700-304-1	Sequence 1, Appl
26	1105.2	44.9	9050	4	US-09-184-418C-7	Sequence 7, Appl
27	1100.4	44.7	7399	2	US-08-418-848A-9	Sequence 9, Appl

398 GCGCCCTGCTGAGCATCAAGTGGCGGCCAGATCAAGGAGGCCCTGCTGACACCGCG 457  
 Db  
 241 GCGCCCTGCTGAGCATCAAGTGGCGGCCAGATCAAGGAGGCCCTGCTGACACCGCG 300  
 Qy  
 458 CCGAGACACCGCTGCTGAGGAGATGAGCTGCCCGGCAAGTGGAGGCCCAAGATGTCG 517  
 Db  
 301 CCGAGACACCGCTGCTGAGGAGATGAACTGCCCGGCAAGTGGAGGCCCAAGATGTCG 360  
 Qy  
 518 GCGGATCGCGCGCTTCATCAAGGTCGCGCAGTACGACCAAGATCTCTGATCGAGATCTCG 577  
 Db  
 361 GCGGATCGCGCGCTTCATCAAGGTCGCGCAGTACGACCAAGATCTCTGATCGAGATCTCG 420  
 Qy  
 578 GCAAGAGCCCATCGCACCGTCTGATCGCGGCCACCGCCGTCGAGCATCATCGCGCGCA 637  
 Db  
 421 GCGCAAGCCCATCGCACCGTCTGATCGCGGCCACCGCCGTCGAGCATCATCGCGCGCA 480  
 Qy  
 638 ACATGCTGACCCAGCTGGGCTGCAACCTGAACTTCCCATCAGCCCGCATCGAGACCGTGC 697  
 Db  
 481 ACTGCTGACCCAGCTGGGCTGCAACCTGAACTTCCCATCAGCCCGCATCGAGACCGTGC 540  
 Qy  
 698 CCGTGAAGCTGAAGCCCGGATGACCGGCCCAAGTGAAGCAGTGGCCCTGACCGGAG 757  
 Db  
 541 CCGTGAAGCTGAAGCCCGGATGACCGGCCCAAGTGAAGCAGTGGCCCTGACCGGAG 600  
 Qy  
 758 AGAAGATCAAGGCCCTGACCGCATCTGCGAGGATGAGAGGAGGCAAGATCAACA 817  
 Db  
 601 AGAAGATCAAGGCCCTGAGTGGATCTGACCGAGTGAAGAGGAGGCAAGATCAACA 660  
 Qy  
 818 AGATCGCCCGGAGAACCCCTCAACAACCCCGTGTGCGCATCAAGAGAGAGGACAGCA 877  
 Db  
 661 AGATCGCCCGGAGAACCCCTCAACAACCCCGTGTGCGCATCAAGAGAGAGGACAGCA 720  
 Qy  
 878 CCAAGTGGCGCAGCTGGTGGATCTCGCGAGCTGAACAGCGCACCCAGGACTTCTGGG 937  
 Db  
 721 CCAAGTGGCGCAGCTGGTGGATCTCGCGAGCTGAACAGCGCACCCAGGACTTCTGGG 780  
 Qy  
 938 AGGTGCAAGCTGGGATCCCGCACCCCGCGCTGAGAGAGAGAGACGTCGCGTC 997  
 Db  
 781 AGGTGCAAGCTGGGATCCCGCACCCCGCGCTGAGAGAGAGAGACGTCGCGTC 840  
 Qy  
 998 TGAAGTGGCGCAGCTGAGTCTTCAAGTGGCGCGCTGAGAGAGAGAGACGTCGCGTC 1057  
 Db  
 841 TGAAGTGGCGCAGCTGAGTCTTCAAGTGGCGCGCTGAGAGAGAGAGACGTCGCGTC 900  
 Qy  
 1058 CATTACCATCCCGAGATCAACAACCGAGATCGTGTATCTACAGGCGCCCGCTGTACGTCG 1117  
 Db  
 901 CATTACCATCCCGAGATCAACAACCGAGATCGTGTATCTACAGGCGCCCGCTGTACGTCG 960  
 Qy  
 1118 TGCCTGAGGCTGAAGGAGCGCCGAGCATCTTCCAGAGCAGATGACCAAGATCTGTCG 1177  
 Db  
 961 TGCCTGAGGCTGAAGGAGCGCCGAGCATCTTCCAGAGCAGATGACCAAGATCTGTCG 1020  
 Qy  
 1178 AGCCCTTCCGCGCCGCAACCCCGAGATCGTGTATCTACAGGCGCCCGCTGTACGTCG 1237  
 Db  
 1021 AGCCCTTCCGCGCCGCAACCCCGAGATCGTGTATCTACAGGCGCCCGCTGTACGTCG 1080  
 Qy  
 1238 GCGACCTGGAGATCGCGCAGACCGCCGAGATCGAGAGCTGCGCAGACCTGCTGC 1297  
 Db  
 1081 GCGACCTGGAGATCGCGCAGACCGCCGAGATCGAGAGCTGCGCAGACCTGCTGC 1140  
 Qy  
 1298 GCTGGGCTTCCACACCCCGCAGAGAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357  
 Db  
 1141 GCTGGGCTTCCACACCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
 Qy  
 1358 GCTAGAGCTGCAACCCCGCAG 1417  
 Db  
 1201 GCTAGAGCTGCAACCCCGCAG 1260  
 Qy  
 1418 GCTAGAGCTGCAACCCCGCAG 1477  
 Db  
 1261 GCTAGAGCTGCAACCCCGCAG 1320  
 Qy  
 1478 ACCCGGCATCAAGGTGCGCGCAGCTGTGTCAAGTGTCTGCGCGCGCCCAAGGCCCTGACCG 1537

1321 ACSCCGGCATCAAGGTGAAGCAGCTGTGCAAGTGTGTCGCGCGCACCAAGGCCCTGACCG 1380  
 Qy  
 1538 ACATCGTGCCTGTGACCGGAGGAGCGAGCTGAGCTGCGCGAGAACCGCGAGATCTCTGC 1597  
 Db  
 1381 AGTGTATCCCTGTGACCGGAGGAGCGAGCTGAGCTGCGCGAGAACCGCGAGATCTCTGA 1440  
 Qy  
 1598 GCGAGCCGCTGACCGCTGTACTAGACCCAGCAAGACCTGCTGGCCGAGATCCAGA 1657  
 Db  
 1441 AGAGCCCGTGCACGAGGTGTACTAGACCCAGCAAGACCTGCTGGCCGAGATCCAGA 1500  
 Qy  
 1658 AGCAGGCCACGACACCTGAGCTGACCTACAGATCTACAGAGGCCCTTCAAGAACCTGAAAG 1717  
 Db  
 1501 AGCAGGCCACGAGGCCAGTGGACCTACAGATCTACAGAGGCCCTTCAAGAACCTGAAAG 1560  
 Qy  
 1718 CCGCAAGTACGCAAGATGCGCACCGCCACACCAAGCAGTGAAGCAGCTGACCGAGG 1777  
 Db  
 1561 CCGCAAGTACGCAAGATGCGCACCGCGCGCCACACCAAGCAGTGAAGCAGCTGACCGAGG 1620  
 Qy  
 1778 CCGTGCAGAAAGATCGCCATGAGAGAGCATCGTGTATCTGGGGCAAGACCCCAAGTTCCGCC 1837  
 Db  
 1621 CCGTGCAGAAAGTGAAGCAGCAGAGAGCATCGTGTATCTGGGGCAAGATCCCAAGTTCAAGC 1680  
 Qy  
 1838 TGCCCATCCAGAAAGAGACCTGCGAGACCTGCTGGACCGACTACTGCGAGGCCACCTGGA 1897  
 Db  
 1681 TGCCCATCCAGAAAGAGACCTGCGAGGCCCTGCTGGATGAGTACTGCGAGGCCACCTGGA 1740  
 Qy  
 1998 TCCCGAGTGGGAGTTCTGTAAACACCCCGCCCTGCTGAACTGTGTATCTGATACAGCTGAG 1957  
 Db  
 1741 TCCCGAGTGGGAGTTCTGTAAACACCCCGCCCTGCTGAACTGTGTATCTGATACAGCTGAG 1800  
 Qy  
 1958 AGGAGCCCATCTGCGCGCGAGACCTTCTACGCTGAGACGCGCGCGCCCAAGCCGAGACCA 2017  
 Db  
 1801 AGGAGCCCATCTGCGCGCGAGACCTTCTACGCTGAGACGCGCGCGCCCAAGCCGAGACCA 1860  
 Qy  
 2018 AGATCGCAAGGCGGCTGCTGACACCGCGCGCGCGCAAGAGATCGTGAAGCTGTGAGCTGACCG 2077  
 Db  
 1861 AGCTGGGAGGCGCGCTGCTGACACCGCGCGCGCGCGCAAGAGTGTGTGAGCATCGCGG 1920  
 Qy  
 2078 AGACCAACCAACCAAGAACCGAGCTGAGCGCATCTGAGCTGGCCCTGCGAGCAGCGGCA 2137  
 Db  
 1921 ACACCAACCAACCAAGAACCGAGCTGAGCGCATCTGAGCTGGCCCTGCGAGCAGCGGCG 1980  
 Qy  
 2138 GCGAGGTGAACATCTGACGACAGCGAGCTAGCGCTGGGCGCATCTCCAGGCCCGAGCGCG 2197  
 Db  
 1981 TGAAGGTGAACATCTGACGACAGCGAGCTAGCGCTGGGCGCATCTCCAGGCCCGAGCGCG 2040  
 Qy  
 2198 ACAAGAGCGAGAGGAGCTGCTGAACCAAGATCTGAGCAGCTGATCAAGAGAGAGAGG 2257  
 Db  
 2041 ACAGAGCGAGAGGAGCTGCTGAGCGAGATCTGAGCGAGATCTCAAGAGAGAGAGG 2100  
 Qy  
 2258 TGTACTGAGCTGGTGGCGCCCGCCAGAGGCGATCGGCGCGCAAGCAGCAGATCGACAAGC 2317  
 Db  
 2101 TGTACTGAGCTGGTGGCGCCCGCCAGAGGCGATCGGCGCGCAAGCAGCAGATCGACAAGC 2160  
 Qy  
 2318 TGGTGAAGCAAGGCGATCGCGCAAGGTGTCTTCTGAGCGGATCGATGCGGCGCATCTGTA 2377  
 Db  
 2161 TGGTGAAGCGCGCATCGCAAGGTGTCTTCTTGAACCGCATCGATGCGGCGCATCTGTA 2220  
 Qy  
 2378 TTTACAGTATCATGAAGACCTGCTGAGCGAGCGCGCGCCCTAGGATCGATTAAAGC 2437  
 Db  
 2221 TTTACAGTATCATGAAGACCTGCTGAGCGAGCGCGCGCCCTAGGATCGATTAAAGC 2280  
 Qy  
 2438 TTTCCCGGGCTAGCACCGGTGAATTC 2463  
 Db  
 2281 TTTCCCGGGCTAGCACCGGTGAATTC 2306

RESULT 2  
 US-09-475-515-84  
 ; Sequence 84, Application US/09475515A  
 ; Patent No.: 6602705  
 ; GENERAL INFORMATION:

; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEDEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE OF INVENTION: OF VIRUS-LIKE PARTICLES  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 84  
; LENGTH: 2312  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: FS(-).protmod.Rtopt(+)  
US-09-475-515-84

Query Match 82.2%; Score 2025.2; DB 4; Length 2312;  
Best Local Similarity 93.3%; Pred. No. 0;  
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;  
3;

Qy 170 GCGCAAGGAGGCGCACAGATGAAGACTGACCGAGGCCAGGCCAAGCTTCTTGGCG 229  
Db 1 GCGCGCGGAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAATTTCTCGCG 60

Qy 230 AGGACCTGGCTTCCCGAGGCAAGCCCGCGAGTTCCCGAGGACAGAACCGGCGCA 289  
Db 61 AGGACCTGGCTTCCCGAGGCAAGCCCGCGAGTTCCCGAGGACAGAACCGGCGCA 120

Qy 290 ACAGCCCCACAGCGCGAGTGCAGTGGCGCGC-----ACAAACCCCGCGAGG 343  
Db 121 ACAGCCCCACAGCGCGAGTGCAGTGGCGCGCGAGAACAAACAGCTGAGCGAGG 180

Qy 344 CCGGCGCGAGCGCGAGGACCTTG-----ACCTTCCCGCGAGTCCCGCTGGCGAGC 397  
Db 181 CCGGCGCGAGCGCGAGGACCTTGAGCTTCAACTTCCCGCGAGTCACTTGGCGAGC 240

Qy 398 GCCCCTGGTGCAGCATCAAGTGGCGCGCGAGATCAAGGAGGCCCTGCTGCACACCGCG 457  
Db 241 GCCCCTGGTGCAGCATCAAGTGGCGCGCGAGATCAAGGAGGCCCTGCTGCACACCGCG 300

Qy 458 CCGACGACACCGTGTGGAGGAGATGAGCTTCCCGCGCAAGTGGAGCCCAAGATGTCG 517  
Db 301 CCGACGACACCGTGTGGAGGAGATGAGCTTCCCGCGCAAGTGGAGCCCAAGATGTCG 360

Qy 518 CGCGCATCGCGGCTTTCATCAAGTGGCGCGCGAGTACGACCGAGTCCGATCGAGATCTCG 577  
Db 361 CGCGCATCGCGGCTTTCATCAAGTGGCGCGCGAGTACGACCGAGTCCGATCGAGATCTCG 420

Qy 578 GCAAGAGGCCATCGGCAACCGTGTGATCGGCGCGCGAGTCCCGCGCAACATCATCGCGCGCA 637  
Db 421 GCAAGAGGCCATCGGCAACCGTGTGATCGGCGCGCGAGTCCCGCGCAACATCATCGCGCGCA 480

Qy 638 ACATGCTGACCGAGTGGGTGACCTTGAATTTCCCGATCAGCCCATCGAGACCGTGC 697  
Db 481 ACCTGCTGACCGAGTGGGTGACCTTGAATTTCCCGATCAGCCCATCGAGACCGTGC 540

Qy 698 CGGTGAAGCTGAAGCCCGCGATGAGCGCGCGAGTGAAGCGTGGCGCGCGAGTCCCGAGG 757  
Db 541 CGGTGAAGCTGAAGCCCGCGATGAGCGCGCGAGTGAAGCGTGGCGCGCGAGTCCCGAGG 600

Qy 758 AGAAGATCAAGCCCTGACCGCCATCTCGGAGGAGATGAGAGGAGGCGCAAGATCACCA 817  
Db 601 AGAAGATCAAGCCCTGAGTGTGAGATCTGACCGAGATGAGAGGAGGCGCAAGATCACCA 660

Qy 818 AGATCGGCGCGGAGAACCCCTTACAAACACCCCGTGTTCGCCATCAAGAGAAAGGACAGCA 877  
Db 661 AGATCGGCGCGGAGAACCCCTTACAAACACCCCGTGTTCGCCATCAAGAGAAAGGACAGCA 720

Qy 878 CCAAGTGGCGCAAGCTGGTGGACTTCGCGAGGTGAACAAAGCGCACCCAGGACTTCTGGG 937  
Db 721 CCAAGTGGCGCAAGCTGGTGGACTTCGCGAGGTGAACAAAGCGCACCCAGGACTTCTGGG 780

Qy 938 AGGTGAGCTGGGATCCCGACCCCGCGCGCTGAAGAAAGAAAGAGCGTGAACCGTGC 997  
Db 781 AGGTGAGCTGGGATCCCGACCCCGCGCGCTGAAGAAAGAAAGAGCGTGAACCGTGC 840

Qy 998 TGGACGTGGGCGAGCCCTACTTTCAGCGTGCCTTGGACGAGGACTTCCGCAAGTACACCG 1057  
Db 841 TGGACGTGGGCGAGCCCTACTTTCAGCGTGCCTTGGACGAGGACTTCCGCAAGTACACCG 900

Qy 1058 CTTTACCATCCCGAGCATCAACAGAGACCCCGCGCATCCGCTACAGTACAAAGTGC 1117  
Db 901 CTTTACCATCCCGAGCATCAACAGAGACCCCGCGCATCCGCTACAGTACAAAGTGC 960

Qy 1118 TGCCCCAGGCTGGAAGGCGACCCCGAGCATTTTCCAGAGAGCATGACCAAGATCTCTGG 1177  
Db 961 TGCCCCAGGCTGGAAGGCGACCCCGCGCATTTTCCAGAGAGCATGACCAAGATCTCTGG 1020

Qy 1178 AGCCCTTCGCGCGCGCAACCCCGAGATCGTGTACTACCA-----GGCCCCCTGTACG 1231  
Db 1021 AGCCCTTCGCGCGCGCAACCCCGAGATCGTGTACTACCA-----GGCCCCCTGTACG 1080

Qy 1232 TGGGCGAGCGACTGAGATCGCGCGAGCACCGCGCGCAAGATCGAGAGCTGCGCAAGCACC 1291  
Db 1081 TGGGCGAGCGACTGAGATCGCGCGAGCACCGCGCGCAAGATCGAGAGCTGCGCGAGCACC 1140

Qy 1292 TGCTCGCTGGGCTTTCACACCCCGCAAGAGACCCAGAGAGCGCCCTTCTCTGT 1351  
Db 1141 TGCTCGCTGGGCTTTCACACCCCGCAAGAGACCCAGAGAGCGCCCTTCTCTGT 1200

Qy 1352 GGATCGGCTAGAGTGCACCCCGAGAGTGCAGCTGCAGCCCATCGAGTGCGCCGAGA 1411  
Db 1201 GGATCGGCTAGAGTGCACCCCGAGAGTGCAGCTGCAGCCCATCGAGTGCGCCGAGA 1260

Qy 1412 AGGAGAGCTGGAACCGTGAACGACATCCAGAAAGTGTGGGCAAGCTGAACTTGGGCGAGCC 1471  
Db 1261 AGGAGAGCTGGAACCGTGAACGACATCCAGAAAGTGTGGGCAAGCTGAACTTGGGCGAGCC 1320

Qy 1472 AGATCTACCCCGCATCAAGTGGCGCGAGCTGCAAGCTGCTGGCGCGCGCAAGGCGCC 1531  
Db 1321 AGATCTACCCCGCATCAAGTGGCGCGAGCTGCAAGCTGCTGGCGCGCGCAAGGCGCC 1380

Qy 1532 TGACCGACATCGTGCCCTGACCGAGGAGCGAGCTGGAGCTGGCCCGAGAACCGCGAGA 1591  
Db 1381 TGACCGAGTGTATCCCTGACCGAGGAGCGAGCTGGAGCTGGCCCGAGAACCGCGAGA 1440

Qy 1592 TCTTGGCGAGCCCGTGCACCGCGTGTACTACGACCCCGAGAGACCTTGTGGCGCGAGA 1651  
Db 1441 TCTTGAAGGAGCCCGTGCACCGAGGTGTACTACGACCCCGAGAGACCTTGTGGCGCGAGA 1500

Qy 1652 TCCAGAGAGAGGCGCACAGCAGTGCAGCTTACAGATCTTACAGAGCGCTTCAAGAAACC 1711  
Db 1501 TCCAGAGAGAGGCGCACAGCAGTGCAGCTTACAGATCTTACAGAGCGCTTCAAGAAACC 1560

Qy 1712 TGAAGACCGCGCAAGTACGCGCAAGATCGCGCACCGCGCGACACCAACGAGTGAAGAGAGTGA 1771  
Db 1561 TGAAGACCGCGCAAGTACGCGCGCATGCGCGCGCGCGCACACCAACGAGTGAAGAGAGTGA 1620

Qy 1772 CCGAGCGCGTGCAGAGATCGCGCATGGAGAGATGTGATCTGGGCGAGAGACCCCAAGT 1831  
Db 1621 CCGAGCGCGTGCAGAGAGTGAAGACCGAGAGATGTGATCTGGGCGAGAGATCTGGGCGAG 1680

Qy 1832 TCCGCTCGCCCATCCAGAAAGGAGACCTGGGAGACCTGGTGGAGCCGAGTACTGGAGGCCA 1891  
Db 1681 TCAAGCTGCCCATCCAGAAAGGAGACCTGGGAGGCGCTGGTGGATGGAGTACTGGAGGCCA 1740

Qy 1892 CTTGGATTCGCCGAGTGGAGTTCGTGAAACACCCCCCTGGTGGAGCTGTGTACAGC 1951

Db 1741 CTTGGATCCCGAGTGGAGTTCTGTAACACACCCCGCTGTTGAAGCTGTGTACACAGC 1800  
Qy 1952 TGGAGAGAGGCCATCATCGGCGCGAGACCTTCTAAGTGAAGCGCGCGCGCAACCGCG 2011  
Db 1801 TGGAGAGAGGCCATCATCGGCGCGAGACCTTCTAAGTGAAGCGCGCGCGCAACCGCG 1860  
Qy 2012 AGACCAAGATCGCAAGGCGGCTACGTGACCGACCGGCGCGCGAGAGATCGTGAGCC 2071  
Db 1861 AGACCAAGATCGCAAGGCGGCTACGTGACCGACCGGCGCGCGAGAGATCGTGAGCC 1920  
Qy 2072 TGACCGAGACCAACCAAGAGACCGAGTGGAGGCCATTCAGCTGGCGCTGCGAGGACA 2131  
Db 1921 TCGCCGACACCAACCAAGAGACCGAGTGGAGGCCATTCAGCTGGCGCTGCGAGGACA 1980  
Qy 2132 GCGGAGCGAGGTGAACATCGTCAACGACAGCGAGTACGCGCTGGGCGATCATCCAGGCC 2191  
Db 1981 GCGGCTGGAGGTGAACATCGTCAACGACAGCGAGTACGCGCTGGGCGATCATCCAGGCC 2040  
Qy 2192 AGCCCGACAAGAGCGAGCGAGTGGTGAACCAAGATCATCGAGCGAGTGTATCAAGAAGG 2251  
Db 2041 AGCCCGACAAGAGCGAGCGAGTGGTGAACCAAGATCATCGAGCGAGTGTATCAAGAAGG 2100  
Qy 2252 AGAGGTGTACCTGAGTGGTGGTCCCGCCCAAGAGGCGATCGGCGCAACGAGCAGATCG 2311  
Db 2101 AGAGGTGTACCTGAGTGGTGGTCCCGCCCAAGAGGCGATCGGCGCAACGAGCAGTGG 2160  
Qy 2312 ACAAGCTGGTGAAGCAAGGCGATCGCAAGTGTGTCTCTGAGCGCGATCGATGGCGGCA 2371  
Db 2161 ACAAGCTGGTGAAGCAAGGCGATCGCAAGTGTGTCTCTGAGCGCGATCGATGGCGGCA 2220  
Qy 2372 TCGTGATCTACCAAGTACATCGAGCGAGTGTACCTGAGCGCGCGCTAGGATCGATT 2431  
Db 2221 TCGTGATCTACCAAGTACATCGAGCGAGTGTACCTGAGCGCGCGCTAGGATCGATT 2280  
Qy 2432 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2463  
Db 2281 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2312

RESULT 3  
US-09-475-515-83  
; Sequence 83, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475.515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 83  
; LENGTH: 2300  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: FS(-) protmod.RTopt.YMMW  
US-09-475-515-83

Query Match 82.0%; Score 2019.2; DB 4; Length 2300;  
Best Local Similarity 93.2%; Pred. No. 0;  
Matches 2150; Conservative 0; Mismatches 139; Indels 18; Gaps 3;

Qy 170 GCGCAAGAGGAGGCCACAGATGAGAGTGCACCGAGCGCGAGCAATCTTCTCCGCG 229  
Db 1 GCGCGCGAGAGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTCCGCG 60  
Qy 230 AGACCTGGGCTTCCCGAGGCAAGCGCGAGTTCCCGAGCGAGAGAACCGCGCCA 289  
Db 61 AGACCTGGGCTTCCCGAGGCAAGCGCGAGTTCCCGAGCGAGAGAACCGCGCCA 120  
Qy 290 ACAGCCCAACAGCGCGAGCTGACGTGCGCGCG-----ACAAACCGCGCGAGG 343  
Db 121 ACAGCCCAACCGCGAGCTGACGTGCGCGCGCGAGAACACAGCTGAGCGAGG 180  
Qy 344 CCGCGCGAGCGCGAGGCAACCTG-----NACTTCCCGAGATCACCTCTGGCAGC 397  
Db 181 CCGCGCGAGCGCGAGGCAACCTGAGCTTAACTTCCCGAGATCACCTCTGGCAGC 240  
Qy 398 GCCCGCTGGTGAAGCATCAAGGTGGCGCGAGATCAAGGAGCGCTGCTGGAACACCGGCG 457  
Db 241 GCCCGCTGGTGAAGCATCAAGGTGGCGCGAGCTCAAGGAGCGCTGCTGGAACACCGGCG 300  
Qy 458 CCGAGGACCGGTGCGAGAGATGAGCTGCGCGCGAGTGAAGCGCGCAAGATGATCG 517  
Db 301 CCGAGGACCGGTGCGAGAGATGAGCTGCGCGCGAGTGAAGCGCGCAAGATGATCG 360  
Qy 518 CCGGATCGGCGGCTTTCATCAAGGTGGCGCGAGTACGACAGATCCCGCTGGAGATCTGCG 577  
Db 361 CCGGATCGGCGGCTTTCATCAAGGTGGCGCGAGTACGACAGATCCCGCTGGAGATCTGCG 420  
Qy 578 GCAAGAGCGCATCGGCAACGCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 637  
Db 421 GCGCAAGAGCGCATCGGCAACGCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 480  
Qy 638 ACATGCTGACCGAGTGGGCTGCAACCTGAACTTCCCGATCAGCGCGCGCGCGCGCGCGTGC 697  
Db 481 ACCTGCTGACCGAGTGGGCTGCAACCTGAACTTCCCGATCAGCGCGCGCGCGCGCGTGC 540  
Qy 698 CCGTGAAGCTGAAGCGCGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 757  
Db 541 CCGTGAAGCTGAAGCGCGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
Qy 758 AGAGATCAAGCGCGCTGACCGCATCTCGAGAGAGATGAGAGAGAGCGCGCGCGCGCGCG 817  
Db 601 AGAGATCAAGCGCGCTGAGAGATCTGACCGAGATGAGAGAGAGCGCGCGCGCGCGCG 660  
Qy 818 AGATCGCGCGCGAGAACCGCTTACAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 877  
Db 661 AGATCGCGCGCGAGAACCGCTTACAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720  
Qy 878 CCAAGTGGCGCAAGCTGGTGAAGTCCGCGAGTGAACAGCGCGCGCGCGCGCGCGCGCG 937  
Db 721 CCAAGTGGCGCAAGCTGGTGAAGTCCGCGAGTGAACAGCGCGCGCGCGCGCGCGCGCG 780  
Qy 938 AGGTGACGTGGGCGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGC 997  
Db 781 AGGTGACGTGGGCGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGC 840  
Qy 998 TGGAGTGGGCGAGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1057  
Db 841 TGGAGTGGGCGAGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
Qy 1058 CCTTCAACCATCCCGAGCATCAACAGCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGTGC 1117  
Db 901 CCTTCAACCATCCCGAGCATCAACAGCGAGACCGCGCGCGCGCGCGCGCGCGCGCGTGC 960  
Qy 1118 TGGCGCGAGGCTGGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGC 1177  
Db 961 TGGCGCGAGGCTGGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGC 1020  
Qy 1178 AGCGCTTCCCGTGC 1237  
Db 1021 AGCGCTTCCCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGC 1080

1238	QY	CGGACCTTGGAGATCGGCCAGCACCGCGCCAAAGATCGAGAGCTGCGAAGCACTGTGTGC	1297
1081	DB	CGGACCTTGGAGATCGGCCAGCACCGCACCAAGATCGAGAGCTGCGCCAGCACTGTGTGC	1140
1298	QY	GCTGGGGCTTCACCAACCCCGACAAGAAGCACACAGAAGAGGCCCCCTTCTCTGTGGATGG	1357
1141	DB	GCTGGGGCTTCACCAACCCCGACAAGAAGCACCAAGAAGGAGCCCCCTTCTCTGCCCCAT--	1198
1358	QY	GCTAGAGCTGCACCCCGACAAGTGGAACCGTGCAGCCCATCGAGCTGCCCGAGAGGAGA	1417
1199	DB	---CGAGCTGCACCCCGACAAGTGGAGCCGTGCGGCCCATCATGCTGTGCCCGAGAAGGACA	1254
1418	QY	GCTGGAACCGTGAACGACATCCAGAAGCTGCTGGGCAAGCTGAACCTTGGGCCCAACGATCT	1477
1255	DB	GCTGGAACCGTGAACGACATCCAGAAGCTGCTGGGCAAGCTGAACCTGAGGCGCAGCATCT	1314
1478	QY	ACCCGGGACATCAAGTTCGCGCCGACTGTGAACTGCTGGCGGGCGCCAAAGCCCTGACCG	1537
1315	DB	ACGCGGGCATCAAGGTGAAGCAGCTGTGAAGCTGCTGGCGGGCAACCAAGGCCCTTGACCG	1374
1538	QY	ACATGCTGCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAACCGCAGATCTCTGC	1597
1375	DB	AGGTATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAACCGCAGATCTCTGA	1434
1598	QY	CGGAGCCGTGCACGGCGTGTACTAGACCCCGACGAAAGAGCTGCTGGCCGAGATCCAGA	1657
1435	DB	AGGAGCCCGTGCAAGAGTGTACTAGACCCCGAAGGAGCTGCTGGCCGCGAGATCCAGA	1494
1658	QY	AGCAGGCGCAGCACCACTGGACCTTACCAGATCTACCAGGAGCCCTTCAAGAACTGTGAAG	1717
1495	DB	AGCAGGCGCAGGCGCAGTGACCTTACCAGATCTACCAGGAGCCCTTCAAGAACTGTGAAG	1554
1718	QY	CGGCAAGTAGCGCAAGACTCGCACCGCCCAACACAAAGAGCTGAAGCAGCTGACCCAGG	1777
1555	DB	CGGCAAGTAGCGCCCGCATCGCGGGCGCCCAACCAACAGCTGAAGCAGCTGACCCAGG	1614
1778	QY	CCGTGCAAGAGATCGCCATGGAGAGCATCGTGATCTTGGGCAAGACCCCAAGTTCGCCC	1837
1615	DB	CCGTGCAAGAGGTGAGCACCGCAGAGCATCGTGATCTTGGGGCAAGATCCCAAGTTCGAAG	1674
1838	QY	TGCCCATCCAGAAGGAGACTGTGGAGACTGTGTGACCGACTACTTGGCAGCGCACTTGG	1897
1675	DB	TGCCCATCCAGAAGGAGACTGTGGAGGCTGTGTGTGATGTGAGTACTGGCAGCGCACTTGG	1734
1898	QY	TCCCGAGTGGAGTTCTGTAAACACCCCCCTCTGTGTAAGCTGTGGTACCAGCTGGAGA	1957
1735	DB	TCCCGAGTGGAGTTCTGTAAACACCCCCCTCTGTGTAAGCTGTGGTACCAGCTGGAGA	1794
1958	QY	AGGAGCCCATCATCGGCGCGGAGACTTCTTACTGTGAACGGCGCCGCAACCGCAGAGCCA	2017
1795	DB	AGGAGCCCATCATCGGCGCGGAGACTTCTTACTGTGAACGGCGCCGCAACCGCAGAGCCA	1854
2018	QY	AGATCGGCAAGCCCGGTACTGTGAACCGAACCGGGCGCGCAGAGAAGTCTGTGAGCCTGACCG	2077
1855	DB	AGCTGGGCAAGGCCCGGTACTGTGAACCGAACCGGGCGCGCAGAGAAGTCTGTGAGCCTGACCG	1914
2078	QY	AGACACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCGAGAGCAGCGGCA	2137
1915	DB	ACACCAACCAACAGAGACCGAGCTGCAGGCCATCCACTTGGCCCTTGCAGAGCAGCGGCC	1974
2138	QY	GGGAGGTGAACATCGTGAACCGACAGCGGAPGACCCCTTGGGCCATCATCCAGGCCAGCCCG	2197
1975	DB	TGGAGGTGAACATCGTGAACCGACAGCGGATACGCCCTTGGGCCATCATCCAGGCCAGCCCG	2034
2198	QY	ACAGAGCGAGAGCGAGCTGTGTGAACAGATCATCATGAGCAGCTGTATCAAGAAGGAGAGG	2257
2035	DB	ACAGAGCGAGAGCGAGCTGTGTGAACAGATCATCATGAGCAGCTGTATCAAGAGAGGAGAGG	2094
2258	QY	TGTTACTAGCTGGTTCGCCGCCCAAGGGCATTCGGCGGCGAAACGAGCAGATTCGACAGC	2317
2095	DB	TGTTACTGGCTGGTTCGCCGCCCAAGGGCATTCGGCGGCGAAACGAGCAGATTCGACAGC	2154
2318	QY	TGGTGAGCAAGGGCATCCGCAAGGTGTGTTCTCTGACGGCATCGATGGCGGCATCTGTGA	2377

Db	2155	TGTTGAGCGCGGCATCCGCAAGGTGCTTCTCTGAAACGGATCGATGGCGGCATCGTGA	2214
Qy	2378	TCTACCACTACATGAGACGACCTGTACGTGGGAGCGGGGGCCCTAGGATCGATTAAAGC	2437
Db	2215	TCTACCACTACATGAGACGACCTGTACGTGGGAGCGGGGGCCCTAGGATCGATTAAAGC	2274
Qy	2438	TTCCCGGGGCTAGCACCGGTGAATTC	2463
Db	2275	TTCCCGGGGCTAGCACCGGTGAATTC	2300
RESULT 4			
US-09-475-515-6			
; Sequence 6, Application US/09475515A			
; Patent No. 6602705			
; GENERAL INFORMATION:			
; APPLICANT: BARNETT, Susan			
; APPLICANT: ZUR NEGEDE, Jan			
; APPLICANT: SRIVASTAVA, Indresh			
; APPLICANT: LIAN, Ying			
; APPLICANT: HARTOG, Karin			
; APPLICANT: LIU, Hong			
; APPLICANT: GREER, Catherine			
; APPLICANT: SELBY, Mark			
; APPLICANT: WALKER, Christopher			
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION			
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES			
; FILE REFERENCE: 1621.002			
; CURRENT APPLICATION NUMBER: US/09/475,515A			
; CURRENT FILING DATE: 1999-12-30			
; NUMBER OF SEQ ID NOS: 90			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 6			
; LENGTH: 4319			
; TYPE: DNA			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence: synthetic			
; OTHER INFORMATION: HIV-Gag-polymerase			
US-09-475-515-6			

Query Match	78.9%;	Score 1942.6;	DB 4;	Length 4319;
Best Local Similarity	89.7%;	Pred. No. 0;		
Matches 2129;	Conservative 0;	Mismatches 225;	Indels 19;	Gaps 4;
Qy 33	GGCCACAGCGCCCAACATCTCTGATGCAGCGCAGCACATTTCAAGGGGCCCCAAGCGCATCAT 92			
Db 1122	GAGCAACTCGCGGACCATCATGATGCAGCGCGGCACATTCGCGAACAGCGAGAACCGGT 1181			
Qy 93	CAAGTGCCTTCAACTGCGGCAAGAGAGGCCACATCGCGCCGCAACTCGCGCGGCCCGCCGCAA 152			
Db 1182	CAAGTGCCTTCAACTGCGGCAAGAGAGGCCACACCGCCAGGAACCTCGCGCGGCCCGCCGCAA 1241			
Qy 153	GAAAGGCTCTTGGAGATGTCGGCAAGAGGGCCACACAGATGAAGGACTGCAACGAGCGCCA 212			
Db 1242	GAAAGGCTCTTGGCGCTCTGGCGCGCGAAGGACACCAAAATGAAAGATGTGCATGTGAGAGACA 1301			
Qy 213	GGCCAACTTCTTCGCGAGAGACCTGGCTTCCGCCAGGCGCAAGGGCCGCGAGTTTCCCCAG 272			
Db 1302	GGCTAA-TTTTITTAGGNAAGATCTGGCTTCTTCAAGGGAGGCGCAGGAAITTTCTTC 1360			
Qy 273	CGAGCAGAAACCGCGCCAAAGCCCCACACGCGCCGAGCTGCAGGTCGCGCGCG-----A 326			
Db 1361	AGAGCAGACCAAGCCACACAGCCCCACACGAAGAGAGCTTTCAGTTTGGGGAGGAGAAA 1420			
Qy 327	CAACCCCGCAGCAGGGCGCGCGCCGAGCGCCAGGGCA-----CCCTGAACTTCCCCCA 380			
Db 1421	CAACTCCCTCTCAGAACGAGGCGCGATAGCAAGGAACGTGATCCTTTAACTTCCCTCA 1480			
Qy 381	GATACCCCTGTGGCAGCGCCCCCTGTGTAGCATCAAGGTGGCGGCCAGATCAAGGAGGC 440			
Db 1481	GATCACTCTTTGGCAAAGACCCCTCTGTCACATGAAGGATCGCGCGCCAGCTCAAGGAGGC 1540			

441 CCTGTGACACCGCGCCGACGACACCGTCTGAGAGATGAGCTGCGCGGCAAGTG 500  
1541 GCTGCTGACACCGCGCCGACGACACCGTCTGAGAGATGAGCTGCGCGGCAAGTG 1600  
501 GAAGCCCAAGATGATCGCGGCGATCGCGGCTTCAACAGGTGCGCGAGTACGACAGAT 560  
1601 GAAGCCCAAGATGATCGCGGCGATCGCGGCTTCAACAGGTGCGCGAGTACGACAGAT 1660  
561 CTTGATCGAGATCTGCGGCAAGAGGCGCATCGGACCGTGTGATTCGCGCCCGCCCGCT 620  
1661 CCGCGTGGAGATCTGCGGCGACAGGCGCATCGGACCGTGTGATTCGCGCCCGCCCGCT 1720  
621 GAACATCATCGCGGCGATCTGCGGCGATCGGACCGTGTGATTCGCGCCCGCCCGCT 680  
1721 GAACATCATCGCGGCGATCTGCGGCGATCGGACCGTGTGATTCGCGCCCGCCCGCT 1780  
681 CCGCATCGAGACCGTGGCGGCGATGAGCGCGGCGATGAGCGCGCGCGCGCGCGCGCG 740  
1781 CCGCATCGAGACCGTGGCGGCGATGAGCGCGGCGATGAGCGCGCGCGCGCGCGCGCG 1840  
741 GTGGCGCGTGGAGAGAGATCAAGCGCGCGATGAGCGCGCGCGCGCGCGCGCGCGCG 800  
1841 GTGGCGCGTGGAGAGAGATCAAGCGCGCGATGAGCGCGCGCGCGCGCGCGCGCGCG 1900  
801 GGAGGCAAGATCAACAGATCGGCGCGCGATGAGCGCGCGCGCGCGCGCGCGCGCGCG 860  
1901 GGAGGCAAGATCAACAGATCGGCGCGCGATGAGCGCGCGCGCGCGCGCGCGCGCGCG 1960  
861 CAAGAAGAGGAGAGACCAAGTGGCGGCGAGTGGTGGATTCGCGAGTGAACAAGCG 920  
1961 CAAGAAGAGGAGAGACCAAGTGGCGGCGAGTGGTGGATTCGCGAGTGAACAAGCG 2020  
921 CACCCAGGAGTCTGCGGAGGTGAGCTGGGCGATCCCGCGCGCGCGCGCGCGCGCGCGCG 980  
2021 CACCCAGGAGTCTGCGGAGGTGAGCTGGGCGATCCCGCGCGCGCGCGCGCGCGCGCGCG 2080  
981 GAGAGCGTACCGTGGAGTGGCGGCGAGTGGTGGATTCGCGAGTGGCGGCGAGGAG 1040  
2081 GAGAGCGTACCGTGGAGTGGCGGCGAGTGGTGGATTCGCGAGTGGCGGCGAGGAG 2140  
1041 CTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAGAGACCGCGCGCGCGCGCG 1100  
2141 CTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAGAGACCGCGCGCGCGCGCG 2200  
1101 CTACCAAGTACACCGCTTCAACATCCCGAGCATCAACAGAGACCGCGCGCGCGCGCGCG 1160  
2201 CTACCAAGTACACCGCTTCAACATCCCGAGCATCAACAGAGACCGCGCGCGCGCGCGCG 2260  
1161 CATGACCAAGATCTGAGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1217  
2261 CATGACCAAGATCTGAGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2320  
1218 ---GGCG 1274  
2321 CATGACCAAGATCTGAGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2380  
1275 GGAGTGGGAGACCGTGGTGGCGGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1334  
2381 GGAGTGGGAGACCGTGGTGGCGGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2440  
1335 GGAGCG 1394  
2441 GGAGCG 2500  
1395 CATGAGTGGCGGAGAGAGAGTGGAGCGCGTGAACAGCATCCAGAGAGTGGTGGGCGAA 1454  
2501 CATGAGTGGCGGAGAGAGAGTGGAGCGCGTGAACAGCATCCAGAGAGTGGTGGGCGAA 2560  
1455 GCTGAATGGCGGCGAGAGATCTACCGCGCGCATCAAGTGGCGCGCGCGCGCGCGCGCGCT 1514  
2561 GCTGAATGGCGGCGAGAGATCTACCGCGCGCATCAAGTGGAGAGAGTGGTGGGCGAGT 2620

1515 GCT 1574  
2621 GCT 2680  
1575 GCGCGAGAACCGCGAGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1634  
2681 GCGCGAGAACCGCGAGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2740  
1635 GCGACCTGGTGGCGGAGATCCAGAGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1694  
2741 GCGACCTGGTGGCGGAGATCCAGAGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2800  
1695 GCGACCTGGTGGCGGAGATCCAGAGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1754  
2801 GCGACCTGGTGGCGGAGATCCAGAGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2860  
1755 GCGACCTGGTGGCGGAGATCCAGAGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1814  
2861 GCGACCTGGTGGCGGAGATCCAGAGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2920  
1815 GCGACCTGGTGGCGGAGATCCAGAGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1874  
2921 GCGACCTGGTGGCGGAGATCCAGAGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2980  
1875 GCGACCTGGTGGCGGAGATCCAGAGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1934  
2981 GCGACCTGGTGGCGGAGATCCAGAGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3040  
1935 GCGACCTGGTGGCGGAGATCCAGAGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1994  
3041 GCGACCTGGTGGCGGAGATCCAGAGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3100  
1995 GCG 2054  
3101 GCG 3160  
2055 GCGAGAGATCGTGGAGCGTGAACCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2114  
3161 GCGAGAGATCGTGGAGCGTGAACCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3220  
2115 GCG 2174  
3221 GCG 3280  
2175 GCGCGATCATCG 2234  
3281 GCGCGATCATCG 3340  
2235 GCGCGTGGTGAAGAGAGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2294  
3341 GCGCGTGGTGAAGAGAGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3400  
2295 GCGCGAACGAGCAGATCGAGAGTGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2354  
3401 GCGCGAACGAGCAGATCGAGAGTGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3460  
2355 GCGCGATCATCG 2387  
3461 GCGCGATCATCG 3493

RESULT 5  
US-09-475-515-80  
; Sequence 80, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine

```

; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 80
; LENGTH: 2305
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(+).Proinact.Rtopt.YM
; US-09-475-515-80

Query Match      76.3%; Score 1878.8; DB 4; Length 2305;
Best Local Similarity 89.4%; Pred. No. 2.5e-306;
Matches 2061; Conservative 0; Mismatches 232; Indels 13; Gaps 3;

Qy 170 GCGCAAGAGGGCCACCATGAAGACTGACACCGAGCCGACGACCACTTCTCCGCG 229
Db 1 GCGGCGCGAAGGACACCAATGAAGATTGCATGAGACAGGCTAATTTTTT-AGGG 59

Qy 230 AGGACCTGGCTTCCCGGCAAGCGCCGAGTTTCCCGAGCAGCAGACCGCGCCA 289
Db 60 AAGATCTGGCTTCTACAGGAGGACGACAGGAAATTTCTTCAGAGCAGACGAGACCA 119

Qy 290 ACAGCCCCACAGCGCGAGCTGAGGTGGCGCGG-----ACACCCCCCGCAGAG 343
Db 120 ACAGCCCCACAGAGAGAGCTTCAAGTTTGGGAGGAGAAACAACTCCCTCTCAGAAG 179

Qy 344 CCGCGCGCGAGCGCCAGGCA-----CCCTGAACCTTCCCGCAGATCACCTCTGGCAGC 397
Db 180 CAGAGCGGATAGACAGGAACTGTATCTTAACTTCCCTCAGATCACTTTTGGCAAC 239

Qy 398 GCCCCTCTGGTAGATCAAGGTGGCGGCGCAGATCAAGAGGCGCTCTGGACCCGCG 457
Db 240 GACCCCTCGTCAATAAAGATCGGGGGCAACTCAAGGAGCGCTGCTCENATACAGGAG 299

Qy 458 CCGAGCACCGCTCTGGAGAGATGAGCTGCGCGCAAGTGGAGGACCAAGATATCG 517
Db 300 CAGATGATACAGTATTAGAGAAATGAATTTGCCAGGAAATGGAACCAAAATGATAG 359

Qy 518 CCGGCATCGCGGCTTCATCAAGTGGCCAGTACGACAGATCTCTGATCGAGATTCGCG 577
Db 360 GGGGATCGGGGCTTCATCAAGTGGAGGAGTACGACAGATACCTGTAGAAATCTGTG 419

Qy 578 GCAAGAGCGCATCGGCAAGCTGTATCGGCCCGCCAGCGTGAACATCATCGCGCGCA 637
Db 420 GACATAAAGCTATAGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA 479

Qy 638 ACATGCTGACCCAGCTGGGCTGCACCTTGAACCTTCCCGCATCAGCCCCATCGAGACCGTGC 697
Db 480 ATCTGTTGACCCAGATCGGCTGACCTTGAATTTCCCGCATCAGCCCTATTGAGACGCTGC 539

Qy 698 CCGTGAAGCTGAAGCCCGGATGAGCGGCCCAAGGTGAAGCATGGCCCTGACCGAGG 757
Db 540 CCGTGAAGTGAAGCGGGGATGAGCGGCCCAAGGTCAAGCAATGGCCATTGACCGAGG 599

Qy 758 AGAGCATCAAGCCCTGACCCCATCTCGAGGAGATGAGAGGAGGAGGCAAGATCACCA 817
Db 600 AGAATCAAGCCCTGTGTGAGATCTGCAACGAGATGAGAGGAGGAGGCAAGATCAGCA 659

Qy 818 AGATCGGCGCCGAGAACCCCTACACACCCCGCTGTTCGCCATCAAGAAGAGAGACGCA 877
Db 660 AGATCGGCGCCGAGAACCCCTACACACCCCGCTGTTCGCCATCAAGAAGAGAGACGCA 719

Qy 878 CCAAGTGGCGCAAGCTGTGTGATCTTCGCGAGCTGAAACAGGCGACCCAGGACTTCTGGG 937
Db 720 CCAAGTGGCGCAAGCTGTGTGATCTTCGCGAGCTGAAACAGGCGACCCAGGACTTCTGGG 779

```

```

Qy 938 AGGTGAGCTGGGATCCCGCCACCGCGCGGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 997
Db 780 AGGTGAGCTGGGATCCCGCCACCGCGCGGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 939

Qy 998 TGAAGTGGGCGAGCGCTTACTTCAAGGTGGCGGCTGACGAGGACTTCCGCAAGTACACCG 1057
Db 840 TGAAGTGGGCGAGCGCTTACTTCAAGGTGGCGGCTGACGAGGACTTCCGCAAGTACACCG 899

Qy 1058 CTTTCAACATCCCGCAGCATCAACAAAGAGACCGCGCGCATCCGCTACCAAGTCAACAGTGC 1117
Db 900 CTTTCAACATCCCGCAGCATCAACAAAGAGACCGCGCGCATCCGCTACCAAGTCAACAGTGC 959

Qy 1118 TGCCCCAGGCGTGAAGGAGCGCCCGCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGG 1177
Db 960 TGCCCCAGGCGTGAAGGAGCGCCCGCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGG 1019

Qy 1178 AGCCCTTCCGCGCGCCGCAACCGCGAGATCGTGTATCTACCAAGGCGCCCGCTGTACGTGGCA 1237
Db 1020 AGCCCTTCCGCGCGCCGCAACCGCGAGATCGTGTATCTACCAAGGCGCCCGCTGTACGTGGCA 1079

Qy 1238 GCGACCTGGAGATCGGCGAGCACCGCGCAGATCGAGGAGCTGCGCAGCAGCACTGCTGC 1297
Db 1080 GCGACCTGGAGATCGGCGAGCACCGCGCAGATCGAGGAGCTGCGCAGCAGCACTGCTGC 1139

Qy 1298 GCTGGGCTTCAACACCGCGCAGCAAGAGCAGCAAGAGAGGAGCGCCCGCTTCTGTGGATGG 1357
Db 1140 GCTGGGCTTCAACACCGCGCAGCAAGAGCAGCAAGAGAGGAGCGCCCGCTTCTGTGGATGG 1199

Qy 1358 GCTACGAGCTGCACCGCGCAGCAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGA 1417
Db 1200 GCTACGAGCTGCACCGCGCAGCAAGTGGACCGTGCAGCCCATCATGCTGCCGAGAGAGAGA 1259

Qy 1418 GCTGAGCGCTGAACGACATCCAGAGCTGTGGGCAAGCTGGAACCTGGGCGCAGCAGATCT 1477
Db 1260 GCTGAGCGCTGAACGACATCCAGAGCTGTGGGCAAGCTGGAACCTGGGCGCAGCAGATCT 1319

Qy 1478 ACCCGCGCATCAAGGTGGCGCGCAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCG 1537
Db 1320 ACCCGCGCATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCG 1379

Qy 1538 ACATCGTGGCTGACCGAGGCGCGAGCTGAGAGCTGGCGGAGAACCGCGAGATCTCTGC 1597
Db 1380 AGGTGATCCCTGACCGAGGCGCGAGCTGAGAGCTGGCGGAGAACCGCGAGATCTCTGA 1439

Qy 1598 GCGAGCGGTGACCGCGCTGTACTACGACCCCGCAGCAGGAGCTGTGTGCGCGAGATCCAGA 1657
Db 1440 AGAGCGCGTGCACGAGGTGTACTACGACCCCGCAGCAGGAGCTGTGTGCGCGAGATCCAGA 1499

Qy 1658 AGCAGGCGCACGACAGTGGACCTTACAGATCTTACAGAGAGCTGTGTGCGCGAGATCCAGA 1717
Db 1500 AGCAGGCGCACGAGCTGGACCTTACAGATCTTACAGAGAGCTGTGTGCGCGAGATCCAGA 1559

Qy 1718 CCGGCAAGTACCGCAAGTGGCGACCGCGCACCGCGCACCGCGCAAGAGAGCTGACCGAGG 1777
Db 1560 CCGGCAAGTACCGCGCATGGCGCGCGCGCGCACCGCGCACCGCGCAAGAGAGCTGACCGAGG 1619

Qy 1778 CCGTGAAGATCGCCATGAGAGCATGTGTATCTGGGCGCAAGACCCCGAAGTTCCGCG 1837
Db 1620 CCGTGAAGATCGCCATGAGAGCATGTGTATCTGGGCGCAAGATCCCGAAGTTCAAGC 1679

Qy 1838 TGCCCATCCAGAGAGAGACTCTGGAGACTGTGTGACCGAGCTGCGGAGGCGCACTGGA 1897
Db 1680 TGCCCATCCAGAGAGAGACTCTGGAGAGCTGTGTGATGAGTGTGTGAGGCGCACTGGA 1739

Qy 1898 TCCCGGAGTGGGAGTTTCGTAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1957
Db 1740 TCCCGGAGTGGGAGTTTCGTAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1799

Qy 1958 AGGAGCGCATTCATCGCGCGCGAGACTTCTTACGTGAAGCTGTGTGTAACAGCTGGAGA 1957
Db 1800 AGGAGCGCATTCGTGGGCGCGAGAGCTTCTTACGTGAAGCTGTGTGTAACAGCTGGAGA 1859

```





1358 GCTACGAGCTGACCCGACAGTGGACCGTGCAGCCCATCGAGTCCCGCAGAGGAGA 1417  
1198 ----CGAGCTGCACCCCGACAGTGGACCGTGCAGCCCATCATGCTGCCGAGAGGACA 1253  
1418 GCTGACCGCTGAACGACATCCAGAGAGCTGGTGGGCAAGCTGAACCTGGGCCAGCCAGATCT 1477  
1254 GCTGACCGCTGAACGACATCCAGAGAGCTGGTGGGCAAGCTGAACCTGGGCCAGCCAGATCT 1313  
1478 ACCCGGCAATCAAGTGGCCGACCTGTGCAAGCTGCTGGCGGCGCCAGGCGCTGACCG 1537  
1314 AGCGCGGCAATCAAGTGGGCAAGAGCTGTGCAAGCTGCTGGCGGCGCCAGGCGCTGACCG 1373  
1538 ACATCGTCCCTGACCCGAGGAGCGGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGC 1597  
1374 AGGTGATCCCTGACCCGAGGAGCGGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGA 1433  
1598 GCGAGCCCGTGCACCGCGTGTACTAGCACCCGACGAAAGAACCTGTGTGGCCGAGATCCAGA 1657  
1434 AGGAGCCCGTGCACCGAGGTGTACTAGCACCCGACGAAAGAACCTGTGTGGCCGAGATCCAGA 1493  
1658 AGCAGGGCCACGACCACTGAGTGTACCATCTTACCAGAGCCCTTCAAGAACCTGAAGA 1717  
1494 AGCAGGGCCACGAGGCGAGTGTACCATCTTACCAGAGCCCTTCAAGAACCTGAAGA 1553  
1718 CCGGCAAGTACGCAAGATGCGCACCGCCCAACCAACGACGTGAAGAGCTGACCGAGG 1777  
1554 CCGGCAAGTACGCGCCGATCGCGGCGCCCAACCAACGACGTGAAGAGCTGACCGAGG 1613  
1778 CCGTGCAGAGATGCGCATGGAGAGATCGTGTCTGGGGCAAGACCCCAAGTTCGCGC 1837  
1614 CCGTGCAGAGATGCGCATGGAGAGATCGTGTCTGGGGCAAGATCCCAAGTTCGAGC 1673  
1838 TGCCCATCCAGAAAGGAGACTGGAGACTGTGGACCGACTTACTGGCAGGCGCACTGGA 1897  
1674 TGCCCATCCAGAAAGGAGACTGGAGAGCTGTGGATGGAGTACTGGCAGGCGCACTGGA 1733  
1898 TCCCGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGTGTGTACCAAGCTGGAGA 1957  
1734 TCCCGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGTGTGTACCAAGCTGGAGA 1793  
1958 AGGAGCCCATCATCGCGCGCGAGACCTTCTAGCTGACGCGCGCCCAACCCGAGACCA 2017  
1794 AGGAGCCCATCGTGGCGCGGAGACCTTCTAGCTGACGCGCGCCCAACCCGAGACCA 1853  
2018 AGATCGCAAGCGCGTACGTGACCGACCGCGGCGCGGAGAGATCTGTAGCTGACCG 2077  
1854 AGCTGGGCAAGGCGCGCTACGTGACCCGACCGGGGCGCGGAGAGTGGTGAAGCTGCGCG 1913  
2078 AGACCAACCAAGAGACCGAGCTGACGAGCATCCAGCTGGCCCTGCAGGACAGCGCA 2137  
1914 ACACCAACCAAGAGACCGAGCTGACGAGCATCCAGCTGGCCCTGCAGGACAGCGCGC 1973  
2138 GCGAGTGAACATCGTGAACCGACAGCCAGTACGCGCTGGGCTGATCCAGGCGCGAGCCG 2197  
1974 TGGAGTGAACATCGTGAACCGACAGCCAGTACGCGCTGGGCTGATCCAGGCGCGAGCCG 2033  
2198 ACAAGAGCAGAGCGAGCTGTGAAACAGATCATCGAGAGCTGTGATCAAGAGGAGAAGG 2257  
2034 ACAAGAGCAGAGCGAGCTGTGAAACAGATCATCGAGAGCTGTGATCAAGAGGAGAAGG 2093  
2258 TGTACCTGAGCTGGTGCCTGCCCAACAGGGCATTCGCGCGCAACGAGAGAGATGCAACAG 2317  
2094 TGTACCTGAGCTGGTGCCTGCCCAACAGGGCATTCGCGCGCAACGAGAGAGTGTGAACAGC 2153  
2318 TGGTGAAGAGGATCCGCAAGTGTCTGTAAGCGCATCGATGGCGGATCGTGA 2377  
2154 TGGTGAAGCGCGGATCCGCAAGTGTCTGTAAGCGCATCGATGGCGGATCGTGA 2213  
2378 TCTACAGTATCATGGAGAGCTGTGATGGGCGAGCGCGCGCCCTAGGATCGATTAAGC 2437  
2214 TCTACAGTATCATGGAGAGCTGTGATGGGCGAGCGCGCGCCCTAGGATCGATTAAGC 2273

2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463  
2274 TTCCCGGGGCTAGCACCGGTGAATTC 2299  
RESULT 7  
US-09-552-950-2  
; Sequence 2, Application US/09552950  
; Patent No. 6541248  
; GENERAL INFORMATION:  
; APPLICANT: Oxford Biomedica (UK) Limited  
; TITLE OF INVENTION: Anti-Viral Vectors  
; FILE REFERENCE: 674524-2004  
; CURRENT APPLICATION NUMBER: US/09/552,950  
; CURRENT FILING DATE: 2000-04-20  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4307  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: gagpol-synGp - codon  
; OTHER INFORMATION: optimised gagpol sequence  
US-09-552-950-2  
Query Match 67.1%; Score 1651.8; DB 4; Length 4307;  
Best Local Similarity 82.5%; Pred. No. 2,7e-268; Indels 19; Gaps 4;  
Matches 1939; Conservative 0; Mismatches 391;  
33 GGCACACAGCGCCAACTCTGATGACGCGCAGCAACTTCAAGGGCCCCAAGCGCATCAT 92  
1110 GACCAACTCCGCTACCATCATGATGACGCGCGCAACTTTCGGAAACCAACGCAAGATCGT 1169  
93 CAAGTGTTCATCTGCGGCAAGAGGCGGCACATGCGCGCAACTGCGCGCCCCCGCAA 152  
1170 CAAGTGTTCATCTGCGGCAAGAGGCGGCACAGCCCCGCAACTGCGAGGCCCCCTAGGAA 1229  
153 GAAGGCTGTGGAAGTGGCGCAAGAGGCGGCACACAGATGAAGATGCAACGAGCGCCA 212  
1230 AAAGGCTGTGGAAGTGGCGCAAGAGGCGGCACACAGATGAAGATGCAACGAGAGACA 1289  
213 GGCCAACTTCTTCGCGAGGACCTTGGCTTCCCGCAGGGAAGCGCGCGAGTCCCCAG 272  
1290 GGTCAA-TTTTTATGGAAGATCTGGCTTCTTACAGGGAAGCGCAGGGAATTTTCTTC 1348  
273 CGACAGAAACCGCGCAACAGCCCGCCAGCGCGCGAGCTGCAGT-----GCGCGCGCA 326  
1349 AGACAGACCGCGCGCAACAGCCCGCCAGCGCGCGAGCTTCAAGTCTGGGGTGGCGCAA 1408  
327 CAACCCCGCGACGAGCG 380  
1409 CAACCTCCCGCTCCGAGCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1468  
391 GATCACCGCTGTGGCAGCG 440  
1469 GGTCAAGCTTTGGCAGCG 1528  
441 CCGTGTGACACCG 500  
1529 TCTCTGACACCG 1588  
501 GAAGCCCAAGATGATTCGGCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 560  
1589 GAAGCCCAAGATGATTCGGCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1648  
561 CCTGATCGAGATCTCGCGCAAGAGGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 620  
1649 CCTCATCGAATCTTCGGCGCAAGAGGCTATCGGTACCGTGTGTGTGGCGCGCGCGCGCG 1708  
621 GAAATCATTCGGCGCGCAACATGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 680  
1709 CAACATCATCGGACGCAACCTGTTGACGCGAGATCGGTGCAAGCTGCTGAACTTCCCCA 1768

QY	681	CCCATTCAGACCGTCCCGTGAAGCTCAAGCCCGCATGACCGCCCAAGGTTGAAGCA	740
Db	1769	CCCTATCAGACCGTACCGTGAAGCTGAAGCCCGGATGACCGCCCGAAGTCAAGCA	1828
QY	741	GTGGCCCTGACCCGAGGAGAAAGATCAAGCCCTGACCGCCATCTGCGAGGAGATGGAGAA	800
Db	1829	ATGGCCATTGACAGAGGAGAAAGATCAAGCCACTGGTGGAGATTGGCACAGAGATGGAAA	1888
QY	801	GGAGGGCAAGATCACCAGATCGGCCCGCAGNACCCCTACACACCCCGCTGTTCCGCAT	860
Db	1889	GGAGGGCAAAATCTCCAGATTGGGCCCTGAGAACCCGTACACACCGCGGTGTTCCGCAAT	1948
QY	861	CAAGAAGAGACAGACACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGAACAAGCG	920
Db	1949	CAAGAAGAGACTCGACGNAATGCGCAAGCTGGTGGACTTCGCGAGCTGAACAAGCG	2008
QY	921	CAACCAGAGCTTCTGGAGGTGACCTGGGCATCCCCACCCCGCGCTGAGAGAGAA	980
Db	2009	CACGCAAGACTTCTGGAGGTTCAGCTGGGCATCCGCAACCCCGAGGCTGAAGAGAA	2068
QY	981	GAAAGAGCTGACCGTGTGACGTGGCGACCGCTTACTTTCAGCGTGCCTTGGACGAGGA	1040
Db	2069	GAAATCCGTGACCGTACTGGATGTGGTGTGATGCTTCTCCGTTCCTGGACGAGA	2128
QY	1041	CTTCGCAAGTACACCGCTTCACCATCCCGAGCATCAACAGAGACCCCGGATCCG	1100
Db	2129	CTTCGCAAGTACACTGCTTCCCAATCCCTTCGATCAACACAGACACCCGGGATTCG	2188
QY	1101	CTACAGTACAACTGCTGCTCCCGAGGCTGGAAGGCGACGCCCATCTTCCAGAGCAG	1160
Db	2189	ATATCATGTAACGCTGTCTCCCGAGGCTGGAAGGCTTCCCGCAATCTCCAGAGTAG	2248
QY	1161	CATGACCAAGATCTTGGAGCCCTTCGCGCCCGCAACCCCGAGATGTGATCTACCA---	1217
Db	2249	CATGACCAAAATCTTGGAGCCCTTCCGCAAAACAGAACCCCGACATCTATCATGTA	2308
QY	1218	---GGCCCCCTGTAGCTGGAGACGACTGGAGATCGGCCAGCAGCCCGCCAGATCGA	1274
Db	2309	CATGATGATCTGTAGCTGGCTCTGATCTAGAGATAGGGCAGACCCGACCAAGATCGA	2368
QY	1275	GGAGCTGCGCAAGACACTCTGCTGGCTGGGGCTTACCAACCCCGCAAGAGACACAGAA	1334
Db	2369	GGAGCTGCGCAGACACTGTGTAGGTGGGAGCTGACCAACCCCGCAAGAGACACAGAA	2428
QY	1335	GGAGCCCCCTTCTGTGATGGCTACAGCTGACACCCCGCAAGTGACCGTGCAGCC	1394
Db	2429	GGAGCCCTCCCTTCTGTGATGGTTACGAGCTGCACCTTGACAAATGGACCTGCAGCC	2488
QY	1395	CATCGAGCTGCCCGAGAGAGAGCTGGACCGTGAAACGACATCCAGAACTGGTGGGAA	1454
Db	2489	TATCGTGTGCCAGAGAAAGACAGCTGGACTGTCAACGACATACAGAACTGGTGGGAA	2548
QY	1455	GCTGAATGGGCCAGCCAGTCTACCCCGCATCAAGGTGCGCCAGCTGTGCAAGTCTCT	1514
Db	2549	GTTGAATGGGCCAGTCAGATTTACCCAGGATTAAGGTGAGGAGCTGTGCAACTCTCT	2608
QY	1515	GCGCGGCCCAAGGCCCTGACCGACATCTGTGCCCTTGACCGAGGCGCGAGCTGAGCT	1574
Db	2609	CCGCGGAACCAAGCACTCACAGAGGTGATCCCTTAAACGAGGCGCGAGCTGAACT	2668
QY	1575	GGCGGAAACCGGAGATCTCTGGCGAGCCCGTGCAGCGGTGTACTAGACCCAGCAAA	1634
Db	2669	GGCAGAAACCGGAGATCTCTAAAGAGCCCGTGCAGCGGTGTACTAGACCCCTCCAA	2728
QY	1635	GGACTGCTGGCCGAGATCCAGAGACGAGGCGCCAGCACTGGAGCTTACAGATCTACCA	1694
Db	2729	GGACTGATCGCCGAGATCCAGAGACGAGGCGCCAGGCACTGACCTATCAGATTTACCA	2788
QY	1695	GGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCCAAGATGCGCACCCGCCACCAAA	1754
Db	2789	GGAGCCCTTCAAGAACTGAGACCGGCAAGTACGCCCGCGATGAGGGTCCCCACACTAA	2848
QY	1755	CGACGTGAAGCAGCTGACCGAGGCGGTGCAAGATCGCCATGGAGAGCATCGTGTATCTG	1814
Db	2849	CGACGTGAAGCAGCTGACCGAGGCGGTGCAAGATCAACCCGAAAGCATCGTGTATCTG	2908
QY	1815	GGGCAAGACCCCAAGTTCCGCTCCCATCCAGAGGAGACCTGGGAGACCTGGTGGAC	1874
Db	2909	GGGCAAGACTCTCTAAGTTCAAGCTGCCCATCCAGAGAAACCTGGGAAACCTGGTGGAC	2968
QY	1875	CGACTACTGGCAGGCGACCTGGATCCCGAGTGGAGTTCGTGAACACCCCCCTGGT	1934
Db	2969	AGAGTATTGGCAGGCGACCTGGATTCTCTGAGTGGAGTTCGTCAACACCCCTCCCTGGT	3028
QY	1935	GAACTGTGTACCAAGCTGGAGAGGCGCCATCATCGCGCCGAGACCTTCTAGCTGGA	1994
Db	3029	GAACTGTGTACCAAGCTGGAGAGGCGCCATCATCTAGTGGCGCCGAAACCTTCTAGCTGGA	3088
QY	1995	CGCGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACGACCGGGCGC	2054
Db	3089	TGGGCGCGCTAACAGGGAGACTAAGCTGGGCAAGCGCGATACGTCACTAACCGGGCGAG	3148
QY	2055	GCAAGATCTGTGAGCTGACCGAGACCAACCAACAGAGACCGAGCTGCAGGCCATCCA	2114
Db	3149	ACAGAAAGTGTGTACCTCTACTGACACCAACCAACAGAGACTGAGCTGCAGGCCATTTA	3208
QY	2115	GCTGGCCCTGCAGGACAGCGGAGGTGAACATCGTGAACCGAGCAGCTACGCCCT	2174
Db	3209	CCTCGCTTTCGAGACTCGCGGCTGGAGGTGAACATCGTGAACAGACTCTCTAGTATGCCCT	3268
QY	2175	GGGCATCATCCAGGCCAGCCCGCAAGAGCGAGCGAGCTGGTGAACCAAGATCATCGA	2234
Db	3269	GGGCATCATTCAGAGCCAGCCAGACAGAGTCCGAGCTGGTTCATCAGATCATCGA	3328
QY	2235	GCAGCTGATCAAGAAAGAGAGGTGTACCTGAGCTGGTGGTGGCCGCCCAAGGGCATCGG	2294
Db	3329	GCAGCTGATCAAGAGGAAAGGTCTATCTGGCTGGGTACCGCCCAACAAAGGCATTGG	3388
QY	2295	CGGCAACGAGCATCGAAGCTGTGAGCAAGGCACTCCGCAAGGTGCTGTTCCTGGA	2354
Db	3389	CGGCAATGAGCAGCTGCACAAAGCTGTCTCGGCTGGCATCAGGAAGTGTATTCTTCTGGA	3448
QY	2355	CGGCATCGA 2363	
Db	3449	TGGCATCGA 3457	
RESULT 8			
US-09-552-950-5			
; Sequence 5, Application US/09552950			
; Patent No. 6541248			
; GENERAL INFORMATION:			
; APPLICANT: Oxford Biomedica (UK) Limited			
; TITLE OF INVENTION: Anti-Viral Vectors			
; FILE REFERENCE: 674524-2004			
; CURRENT APPLICATION NUMBER: US/09/552,950			
; CURRENT FILING DATE: 2000-04-20			
; NUMBER OF SEQ ID NOS: 22			
; SOFTWARE: Patent In Ver. 2.1			
; SEQ ID NO 5			
; LENGTH: 9772			
; TYPE: DNA			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP			
US-09-552-950-5			
Query Match 66.0%; Score 1624.6; DB 4; Length 9772;			
Best Local Similarity 81.8%; Pred. No. 1.1e-263;			
Matches 1922; Conservative 0; Mismatches 408; Indels 19; Gaps 4;			
QY	33	GGCCACCAAGCCCAACATCTCTGATGCGAGCGCAGCAACTTCAAGGGCCCCCAAGCGCATCAT	92
Db	2217	GACCAACTCCGCTACCATCATGATGCGAGCGCGGCACTTTTCGGAAACCAACGCAAGATCGT	2276

QY 93 CAAGTGTCTTCAACTGCGCAAGAGGGCCACATGCGCCGAACCTGCGCGCCCCCGCAA 152  
Db 2277 CAAGTGTCTTCAACTGCGCAAGAGGGCCACATGCGCCGAACCTGCGCGCCCCTAGGAA 2336  
QY 153 GAAGGGCTCTGGAAGTGGCGCAAGAGGGCCACAGATGAGACTGACCGAGCGCA 212  
Db 2337 AAAGGGCTGTGGAAATGTGGAAGAGGAGGACCAACAAATGAAAGATTTGACTGAGAGCA 2396  
QY 213 GGCCTAACTTCTTCCGCGAGGACTGCGCTTCCCTCCAGGGCAAGGCGCGGAGTTCCTCCAG 272  
Db 2397 GGCCTAA-TTTTTTAGGGAAGATCTGGCTTCCCAACAGGGAAGGCGAGGAAATTTCTTC 2455  
QY 273 CGAGCGAAGACCGGCGCAACAGCGCCACAGCGCGAGCTGAGGTGCGGG-----CGA 326  
Db 2456 AGAGCAGACAGAGCCAAACAGCGCCACAGAGAGAGAGCTTCAGGTGTTGGGGAAGAGCAA 2515  
QY 327 CAACCCCGCAGAGCGCGCGCGCGAGCGCCAGGGCA-----CCCTGAACTTCCCGCA 380  
Db 2516 CAACTCCCTCTCAGAGCAGAGGCGCGATAGACAAAGAACTGTATCCTTTAGCTTCCCTCA 2575  
QY 381 GATCACTCTGTGCGAGCGCGCGCTGTGTGAGCATCAAGGTGGGCGCGCAGATCAAGAGGC 440  
Db 2576 GATCACTCTTGTGCGAGCGCGCGCTGTGTGAGCATCAAGGTGGGCGCGCAGCTCAAGAGGC 2635  
QY 441 CTGCTGGAACCGCGCGCGCGAGCACACCGTGTGAGAGAGATGAGCTGCCCGGCAAGTG 500  
Db 2636 TCTCTTGGACCGCGAGCGAGCAGACACCGTGTGAGAGAGATGTGCTTCCAGGCGCGTG 2695  
QY 501 GAAGCCCAAGATGATCGGCGCGCATCGGCGCTTTCATCAAGGTGGCGCGCATGACACAGAT 560  
Db 2696 GAAGCCCAAGATGATCGGCGCGCATCGGCGCTTTCATCAAGGTGGCGCGCATGACACAGAT 2755  
QY 561 CTGTATCGAGATCTGCGGCAAGAGGCGCATCGGCGCTGCTGATCGGCGCGCGCGCGCT 620  
Db 2756 CCTCATCGAATCTGCGGCGCAAGAGGTATCGGTACCGTGTGCTGGGCGCGCGCGCGCT 2815  
QY 621 GAACATCATCGGCGCGCAACATGCTGACCCAGTGGGCTGACCCCTGAACTTCCCATCAG 680  
Db 2816 CAACATCATCGGCGCGCAACCTGTGACAGATCGGTGCGCGCTGAACTTCCCATTAG 2875  
QY 681 CCCATCGAGACCGTGGCGCGAGAGGCGCATCGGCGCTGCTGATCGGCGCGCGCGCGCT 740  
Db 2876 CCTATCGAGCGGTACCGGTGAGGTGAAGCGCGCGGATGAGCGCGCGCGAGGTCAAGCA 2935  
QY 741 GTGCGCCCTCACAGAGAGAGATCAAGGCGCTGACCGCATCTGCGAGAGATGAGAGAA 800  
Db 2936 ATGCGCCATTGACAGAGAGAGATCAAGGCGCTGAGATTTGACAGAGATGAGAGAA 2995  
QY 801 GAGGGCAAGATCAACAGATCGGCGCGAGAGACCGCTCAACACCGCGGTGTTCCCAT 860  
Db 2996 GGAAGGGAATCTTCAAGATTTGGGCGCTGAGAACCCCGTACACACCGCGGTGTTCCCAT 3055  
QY 861 CAAGAGAGAGAGAGAGAGAGTGGCGCAAGCTGTGAGCTTCCGCGAGCTGAAACAGCG 920  
Db 3056 CAAGAGAGAGAGAGAGAGTGGCGCAAGTGGCGCAAGCTGTGAGCTTCCGCGAGCTGAAACAGCG 3115  
QY 921 CACCGAGAGCTTCTGGAGGTGAGTGGGCGATCCCGCAACCGCGCGGCTGAGAGAGAA 980  
Db 3116 CACGCAAGAGCTTCTGGAGGTTCAGCTGGGCGATCCCGCAACCGCGCGGCTGAGAGAGAA 3175  
QY 981 GAAGAGGTGACCGTGTGAGGTGGGCGAGCGCTACTTCAGCGTGGCGCGCTGACAGGAG 1040  
Db 3176 GAAATCGTGAGCGTACTGAGATGGGTGATGCGCTACTTCTCGTTCCTTGGAGCGAGAG 3235  
QY 1041 CTTCCGCAAGTACACCGCTTTCACATCCCGAGCATCAACAAAGAGACCGCGCGGATCCG 1100  
Db 3236 CTTCAGGAGTACACTGCTTTCACATCCCTTCGATCAACAAAGAGACACCGCGGATTCG 3295  
QY 1101 CTAACGATACACCGTGTGCGCGAGGCTGAGAGGCGAGCGCGAGCATCTTCCAGAGCAG 1160  
Db 3296 ATATCAGTACACCGTGTGCGCGAGGCTGAGAGGCTCTCCCGCAATCTTCCAGAGTAG 3355  
QY 1161 CATGACCAAGATCTCTGGAGCGCTTTCGCGCGCGCGCAACCGCGAGATCGTGTACTACCA--- 1217

Db 3356 CATGACCAAAATCTTGAGGCTTTTCGCAAAACAGAACCCCGACATCGTATCATCAGTA 3415  
QY 1218 ---GGCGCCCTCTAGTGGGCGAGCACTGAGATGCGGCCAGCACCGCGCCCAAGATCGA 1274  
Db 3416 CATGGATGACTTGTAGTGGGCTCTGATCTAGAGATAGGGCAGCACCGCAAGATCGA 3475  
QY 1275 GAGCTGCGCAAGCACTGCTGCGCTGGGCTTCAACCCCGCAACAGAGCACCAGAA 1334  
Db 3476 GAGCTGCGCAAGCACTGCTGAGGTGGGACTGACCAACACCGCAACAGAGCACCAGAA 3535  
QY 1335 GAGCGCCCTTCTCTGTGGATGGGCTACGAGCTGCAACCGCAACAGTGAGCGCGTGCAGCC 1394  
Db 3536 GAGCGCTCTCTCTGTGGATGGGTTACGAGCTGCAACCGCAACAGTGAGCGCGTGCAGCC 3595  
QY 1395 CATGAGCTGCGCAAGAGAGAGCTGGACCGTGAAGACATCCAGAGCTGTGGGCAA 1454  
Db 3596 TATGCTGCTGCCAGAGAGAGAGCTGGACTGTCAACGACATACAGAGCTGTGGGAA 3655  
QY 1455 GCTGAATCTGGCGCAGCCAGATCTACCCCGCATCAAGGTGCGCGCTGTGCAAGTGTCT 1514  
Db 3656 GTTGAATCTGGCGCAGTCAAGATTTACCCAGGATTAAGGTGAGGCGCTGTGCAAACTCCT 3715  
QY 1515 GCGCGCGCGCAAGCGCTGACCGACATCTGCGCGCTGCGCGAGGCGCGAGCTGGAGCT 1574  
Db 3716 CCGCGGAACCAAGCACTCAAGAGGTGATCCCGCTAACCGAGGAGCGCGAGCTGCACT 3775  
QY 1575 GCGCGAACAACCGGAGATCTGCGCGAGCGCGCTGCGCGCTGTACTACGACCCCGACAA 1634  
Db 3776 GCGAAGAAACCGAGAGATCTTAAGGAGCGCGCTGCGCGCTGTACTATGACCCCTCAA 3835  
QY 1635 GGACTGTGGTGGCGAGATCCAGAGAGGCGCGAGCGCGAGCTGAGCTTACAGATCTACCA 1694  
Db 3836 GGACTGTATCGCGAGATCCAGAGAGGCGCGAGCGCGAGCTTATCAGATTTACCA 3895  
QY 1695 GGAGCCCTTCAAGAACTGGAAGCGCGCAAGTACCGCAAGATGCGCACCGCCACACCAA 1754  
Db 3896 GGAAGCCCTTCAAGAACTGGAAGCGCGCAAGTACCGCGAGTGAAGGCGCTGCCACACTAA 3955  
QY 1755 CGAGTGAAGCAGCTGACCGAGGCGGTGCGAGAGATCGCATGAGAGCATCTGTGATCTG 1814  
Db 3956 CGAGTCAAGCAGCTGACCGAGGCGGTGCGAGAGATCACCACCGAAAGCATCTGTGATCTG 4015  
QY 1815 GGGCAAGACCCCAAGTTCGCGCTGCGCATCCAGAGAGAGCTGCGAGAGCTTGTGGAG 1874  
Db 4016 GGGAAAGACTCTTAAGTTCAAGCTGCGCATCCAGAGAGAACTTGGGAACTTGTGTGAC 4075  
QY 1875 CGACTACTGCGAGCGCACTTGGATCCCGAGTGGAGTTCGTGAACACCCCGCGCTGCT 1934  
Db 4076 AGAGTATTGGCAGCGCACTTGGATTCCTGAGTGGAGTTCGTCAACACCCCGCTGCT 4135  
QY 1935 GAAGCTGTGTAACGAGTGGAGAGAGCGCATCATCGCGCGAGAGCTTCTACGTGGA 1994  
Db 4136 GAAGCTGTGTAACGAGTGGAGAGAGCGCATCATCGCGCGCGCGCGCGCGCGCGCG 4195  
QY 1995 GCGCGCGCGCGCGCGAGAGAGAGATCGGCAAGCGCGCGCTTACGTGACCGAGCGCGCG 2054  
Db 4196 TGGGCGCGCTTAACGAGGAGACTTAAGTGGGCAAGCGCGATACGTACTAACCGCGCGCG 4255  
QY 2055 GCGAAGATCTGTGAGCTGACCGAGACCAACCAAGAGAGCGCGCTGCGAGCGCATCCA 2114  
Db 4256 ACAGAGGTGTGTCACTCTGACCAACCAAGAGAGCTGAGCTGCGAGGCGCATTTA 4315  
QY 2115 GCTGCGCTGCGAGGAGAGCGCGAGGTGAACATCGTGAACCGAGCGCGAGCTAGCGCCT 2174  
Db 4316 CTTCTGCTTGGAGGACTCGGGCTTGGAGTGAACATCGTGAACAGCTCTCAGTATGCCCT 4375  
QY 2175 GCGCATCATCTCAGGCGCGCGAGCGCGAGAGCGAGCTGTGTAACCGAGCATCGA 2234  
Db 4376 GCGCATCATCTCAGGCGCGCGAGCGCGAGAGTGTGCGAGTGTGTAACATCATCGA 4435  
QY 2235 GAGCTGTATCAAGAGGAGAGCTGTACTGAGTGGGTGCGCGCGCGCGCGAGCGCATCGG 2294

Db 4436 GCAGCTGATCAAGAGGAAAGGTTCTATCTGGCTGGTACCAGCCCAACAAAGGCAATGG 4495  
 QY 2295 CGGCAACGAGCAGATCGACAGCTGGTGGAGCAAGGGGATCCGCAAGGTGCTGTCTCTGA 2354  
 Db 4496 CGGCAATGAGCAGCTCGACAGCTGGTCTGGCTGGGATCAGGAAGTGTCTTCTGGA 4555  
 QY 2355 CGGCAATCA 2363  
 Db 4556 TGGCATCA 4564

RESULT 9  
 US-09-872-733A-6  
 ; Sequence 6, Application US/09872733A  
 ; Patent No. 6656706  
 ; GENERAL INFORMATION:  
 ; APPLICANT: The Government of the United States of America, as  
 ; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND  
 ; FILE OF INVENTION: SIV ENV GENES  
 ; FILE REFERENCE: 2026-4287US1 HIV GAG/POL, SIV GAG & ENV  
 ; CURRENT APPLICATION NUMBER: US/09/872,733A  
 ; CURRENT FILING DATE: 2001-06-01  
 ; PRIOR APPLICATION NUMBER: PCT/US00/34985  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/173,036  
 ; PRIOR FILING DATE: 1999-12-23  
 ; NUMBER OF SEQ ID NOS: 19  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 6  
 ; LENGTH: 8366  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: DNA sequence  
 ; OTHER INFORMATION: of the construct pCMVgagpolenkan containing a CMV  
 ; OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin  
 ; OTHER INFORMATION: resistance gene  
 US-09-872-733A-6

Query Match 53.6%; Score 1566; DB 4; Length 8366;  
 Best Local Similarity 80.2%; Pred. No. 6.7e-254; Mismatches 446; Gaps 4;  
 Matches 1885; Conservative 0; Indels 19; Gaps 4;

QY 33 GGCCACACGCGCCCAACATCTGTGTCAGCGCAGCACTTCAAGGGGCCCAAGCGCATCAT 92  
 Db 1879 GACGAATCTCGCGACCAATATGATGACAGAGGCACTTCCGGAACCAAGCGGAAGATCT 1938  
 QY 93 CAAGTGTCTCACTCGCGCAGAGGCGCCATCGCCGCAACTGCGCGCCCGCCCGCA 152  
 Db 1939 CAAGTGTCTCACTGTGCAAGAGGCGCCATCGCCGCAACTGCGCGCCCGCCCGCA 1998  
 QY 153 GAAGGGCTGTGGAAGTGGCAAGGAGGCGCCACAGATGAAGGACTGCAACCGAGCGCA 212  
 Db 1999 GAAAGGCTGTGGAAGTGGCAAGGAGGCGCCACCAATGAAGATTGTACTGAGAGCA 2058  
 QY 213 GSCCAACTTCTTCGCGAGGACTGCGCTTCCCGCGCAAGGCGCGGAGTTCGCCAG 272  
 Db 2059 GGTCTAA-TTTTATAGGAGATCTGGCTTCTCAAGGAGGCGCAGGAAATTTCTTC 2117  
 QY 273 CGAGCAGAACCGCGCCCAACAGCCCAACAGCGCGAGCTGCAAGTGGCGG-----CGA 326  
 Db 2118 AGAGCAGACCAAGAGCGCAACAGCCCGCCACAGAGAGAGCTTCAGTCTGGGGTAGAGCAA 2177  
 QY 327 CAACCCCGCAGCGGCGCGCGCGCGCGCGCGCA-----CCTGAACCTTCCCGCA 380  
 Db 2178 CAATCTCCCTTCAGACGAGGCGGATAGACAGGAGTGTATCTTAACTTCCCTCA 2237  
 QY 381 GATCACTCTGTGAGAGCGCCCTCTGTGTAGCATCAAGTGGCGCGCGCAGATCAAGAGGC 440  
 Db 2238 GATCACTCTTGGCAACAGCCCTCTGTGTAGCATCAAGTGGCGCGCGCAGATCAAGAGGC 2297  
 QY 441 CCTGTGACACCG 500

Db 2298 GCTGCTCGATACAGGAGCAGATGATACAGTATTAGAGAAATGAGTTTCCAGGAAGATG 2357  
 QY 501 GAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAGGTGCGCCAGTACGACAGAT 560  
 Db 2358 GAAACCAAAATGATAGGGGGGATCGGGGGCTTCAATCAAGGTGAGGCAATGACGACAGAT 2417  
 QY 561 CTTGATCGAGATCTCGGCAAGAGGCCATCGGACCGTGTGATCGGCGCCCGCCCGCT 620  
 Db 2418 ACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGT 2477  
 QY 621 GAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCAACCTGAACTTCCCGCATCAG 680  
 Db 2478 CAACATAATTGGAAGAAATCTGTTGACCCAGATCGGCTGCACCTTGAATCTCCCGCATCAG 2537  
 QY 681 CCCCATCGAGACCGTCCCGTGAAGTGAAGCCCGCATGAGCGCCCGCATGAGCGCCCGCATGAGCGA 740  
 Db 2538 CCTATTGAGACCGTCCCGTGAAGTTGAAGCCCGGATGAGCGCCCGCATGAGCGA 2597  
 QY 741 GTGCGCCCTCACGAGGAGAAATCAAGCCCTGACCGCATCTCGGAGGAGATGAGAA 800  
 Db 2598 ATGCGCATTCAGCAAGAGAAATCAAGCCCTTAGTCCGAATCTGTACAGAGATGAGAA 2657  
 QY 801 GGAGGCGAGATCACCAAGTGGCG 860  
 Db 2658 GGAAGGAGAGATCAGCAAGATCGGCGCTGAGAAACCTTACAACTTCCAGTCTTCCGAT 2717  
 QY 861 CAAGAAAGAGGAGCAGCACCAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 920  
 Db 2718 CAAGAAAGAGGAGCAGTACCAAGTGGAGAAAGTGGTGGACTTACAGAGAGCTGAAACAAG 2777  
 QY 921 CACCGAGCTCTCGGAGGAGTGGCG 980  
 Db 2778 AACTCAGGATCTTGGGAAGTTTCACTGGGCGATCCCAATCCCGTGGTGGAGAGAA 2837  
 QY 981 GAAGAGCGTGAACGCTCTGAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1040  
 Db 2838 GAAATCAGTGAACGCTCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2897  
 QY 1041 CTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAACACAGAGAGCGCGCGCGCGCGCGCGCG 1100  
 Db 2898 CTTCAGGAGTACACTGCTTCCAGTACCTAGTACAAACAGAGACACAGAGATCCG 2957  
 QY 1101 CTACAGTCAACGCTCTGCG 1160  
 Db 2958 CTACAGTCAACGCTCTGCG 3017  
 QY 1161 CATGACCAAGTCTGAGCGCTTCCG 1217  
 Db 3018 CATGACCAAGTCTGAGCGCTTCCG 3077  
 QY 1218 ---GGCG 1274  
 Db 3078 CATGACCAAGTCTGAGCGCTTCCG 3137  
 QY 1275 GGAGCTGCGCAAGCAGCTGCTGCG 1334  
 Db 3138 GGAGCTGAGACAGCTGCTGAGGCTGGGACTGACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3197  
 QY 1335 GGAGCG 1394  
 Db 3198 GGAACCTTCCCTTCTGAGTGGGCTACGAACTGCTGCTGCAAGTGGAGTGGAGTGGAGTGGAGTGG 3257  
 QY 1395 CATGAGCTGCGCAAGAGAGAGTGGAGCTGGAACGACATCCAGAGAGCTGTTGGGCA 1454  
 Db 3258 CATGCTGCTGCTGAGAGAGAGAGCTGGAATCTGGAACGACATACAGAGAGCTGTTGGGCA 3317  
 QY 1455 GCTGAACCTGCGCAGCAGATCTACCG 1514  
 Db 3318 GTTGAACCTGCGCAGCAGATCTACCG 3377  
 QY 1515 GCG 1574  
 Db 3378 TCGAGGAAACCAAGGACTGACAGAGTGTATCCACTGACAGAGAGAGAGAGAGAGTGAAGT 3437



Db 1924 CCTACAACTCCAGTCTTTCGCAATCAAGAAAGACAGTACCAAGTGGAGAAAGCTG 1983  
Qy 895 GTGACATTCGCGAGCTGAACAAAGCGACCCAGAGCTTCTGGAGGTGTCAGCTGGGCATC 954  
Db 1984 GTGACATTCAGAGAGCTGAACAAAGAACTCTAGGACTTCTGGAGTTCAGCTGGGCATC 2043  
Qy 955 CCCACCCCGCGGCTGAGAAAGAAAGAAAGAGCTGACCGTGTGAGAGTGGGCGAGCC 1014  
Db 2044 CCACATCCCGTGGTGTGAAGAAAGAAAGTCACTGACAGTGTGGATGTGGGTGATGCC 2103  
Qy 1015 TACTTCACGCTGCCCTGGAGAGGACTTCGCAAGTACACCGGCTTTCACCATCCCGAGC 1074  
Db 2104 TACTTCGCTTCCCTTGGACGAGACTTCAGGAGTACACTGCTTCAAGTACCTAGC 2163  
Qy 1075 ATCAAAACAGAGACCCCGGATTCGCTACAGTACAAAGTGTGCTGCCAGGCTGGAAG 1134  
Db 2164 ATCAAAACAGAGACCCAGGATTCGCTACAGTACAAAGTGTGCTGCCAGGATGGAAG 2223  
Qy 1135 GCGAGCCCGAGCTTCCAGAGAGCATGACCAAGATCTCGAGCCCTTCCGCGCCCGC 1194  
Db 2224 GATCACAGCCATCTTTCAGAGAGCATGACCAAGATCTCGAGCCCTTCCGCAAGCA 2283  
Qy 1195 AACCCCGAGATCGTGAATCTACCA-----GGCCCCCTGTAGTGGGCGAGCGACTGGAG 1248  
Db 2284 AACCCAGACATCGTGAATCTATCAGTACATGAGACGACTCTAGTAGGAAAGTACCTGGAG 2343  
Qy 1249 ATC-GGCCAGACCGCGCCAGATCGAGAGCTGCGCAAGCACTGTGCGCTGGGGCTT 1307  
Db 2344 ATCGGGGAGCACAGGACCAAGATCGAGAGCTGAGACAGCATCTGTGAGGTGGGACT 2403  
Qy 1308 CACACCCCGGACCAAGAACCCAGAGAGGAGCCCGCTTCTGTGGATGGGTACGAGCT 1367  
Db 2404 GACCAACCCAGAACAGAACCCAGAGGAACTCCCTTCTGTGGATGGGTACGAACT 2463  
Qy 1368 GCACCCGAGAGTGAACCGTGCAGCCATCCAGCTGCCGAGAGAGGAGCTGACCGT 1427  
Db 2464 GCATCTGACAACTGACAGTGCAGCCCATCTGTGCTGTGAGAGGACAGCTGACTGT 2523  
Qy 1428 GAACGACATCCAGAGCTGTGGGCAAGCTGAACCTGGGCGACCGAGTACCCCGGAT 1487  
Db 2524 GAACGACATACAGAGCTGTGGGCAAGTTGAACCTGGGCAAGCAGATCTACCCAGGAT 2583  
Qy 1488 CAAGTGGCGGACCTGTGACAGTGTGCGCGCGCGAGGCGCTGACCGACATCGTGC 1547  
Db 2584 CAAAGTTAGGACCTGTGAAGCTGTCTGAGGAACTGAGGCACTGACAGAGTATCCC 2643  
Qy 1548 CTTGACCGAGGCGGAGCTGAGCTGGCGAGAACCGCGAGATCTCTCGCGAGCCGT 1607  
Db 2644 ACTGACAGAGAGAGAGCTGAACTGGCAGAGAACCGAGAGATCTTGAAGAGCCAGT 2703  
Qy 1608 GCACGGGTGTACTACGCCAGAGAGCTGTGGTGGCGAGATCCAGAGCAGGCGCA 1667  
Db 2704 ACATGGAGTGTACTACGACCCAGCAGGACCTGTGCGAGAGATCCAGAGCAGGGCA 2763  
Qy 1668 CGACCGTGGACCTTACAGATCTACAGAGGAGCCCTTCAAGAACTGAAAGACCGGCAAGTA 1727  
Db 2764 AGGCAATGAGACCTTACCAATCTACAGAGAGCCCTTCAAGAACTGAGAGCAGGCAAGTA 2823  
Qy 1728 CGCAAGATGCGACCGCCACCAACAGAGCTGAGAGCTGACGAGCCGTGAGAA 1787  
Db 2824 CGCAAGATGAGGGGTGCCACACCAACAGAGTGTGAAGCAGCTGACAGAGCAGTGCAGAA 2883  
Qy 1788 GATCGCATGAGAGCATCTGTGCTGGGCAAGACCCCAAGTTCGCTGCGCATCCA 1847  
Db 2884 GATCACACAGAGAGCATCTGTGCTGGGCAAGACTCCCAAGTTCAGCTGCGCATACA 2943  
Qy 1848 GAAGGAGACTGGAGACCTGTGTGAACCGACTACTGCGAGGCGCACTGTGATTCGCCAGTG 1907  
Db 2944 GAAGGAGACATGGGAGACATGTGTGAACCGAGTACTGGCAAGCCACTGTGATTCGCTGAGTG 3003  
Qy 1908 GGAGTGTGTGAACACCCCGCTGTGTGAAGCTGTGTGATCTGAGTGGAGAGAGGAGCCAT 1967  
Db 3004 GGAGTGTGTGAACACCCCGCTTCCCTTGGTGAAGTGTGTGATCTGAGTGGAGAGGAGCCAT 3063

Qy 1968 CATCGGCGCCGAGACCTTCTACGTGGAGCGGCGCGCAACCGCGAGACCAAGATCGGCAA 2027  
Db 3064 COTGGAGACAGAGACCTTCTACGTGGATGGGCGAGCCCAACAGGAGACCAAGCTGGGCAA 3123  
Qy 2028 GCGCGGCTACGTGACCGACCGGCGCGGCAAGATGTGTGAGCTGACCGAGACCAACCA 2087  
Db 3124 GCGAGGCTACGTGACCAACCGGAGGACGACAGAAAGTGTGTGACCTGACTGACACCCAA 3183  
Qy 2088 CCAGAGACCGAGCTGACGAGCCATCCAGCTGGCGCTGCGAGACAGCGGACGAGGTGAA 2147  
Db 3184 CCAGAGACTGAGCTGACAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGA 3243  
Qy 2148 CATCGTACCGACAGCCAGTACGCTGGGCGATCATCCAGGCCCGAGCCGAGAGGGA 2207  
Db 3244 CATCGTACAGACTCACAGTACG-CATGGGCTATCATCCAGCACAACAGACCAATCCGA 3302  
Qy 2208 GAGCGAGCTGTGAACAGCATCATCGAGCAGCTGATCAAGAGGAGAGTGTACTCTGAG 2267  
Db 3303 GTAGAGCTGGTGAACAGCATCATCGAGCAGCTGATCAAGAGGAGAAAGTGTACTTGC 3362  
Qy 2268 CTGGGTGCGCGCCCAAGAGGATCGCGGCAACAGAGCAGATCGACAGCTGGTGAAGCA 2327  
Db 3363 ATGGGTACCGACACACAAAGGAATTTGGAGAAATGAACAAGTAGATAAATAGTCAGTGC 3422  
Qy 2328 GGCATCCGAGAGTCTCTTCTCGAGCGCATCGAT 2364  
Db 3423 TGGGATCCGAAAGGTCTGTCTCTCGAGCGGATCGAT 3459

RESULT 11  
US-09-184-418C-8  
; Sequence 8, Application US/09184418C  
; Patent No. 6492110  
; GENERAL INFORMATION:  
; APPLICANT: Hahn, Beatrice  
; APPLICANT: Gao, Feng  
; APPLICANT: Shaw, George  
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN  
; FILE REFERENCE: D6287  
; CURRENT APPLICATION NUMBER: US/09/184,418C  
; CURRENT FILING DATE: 1999-11-02  
; NUMBER OF SEQ ID NOS: 112  
; SEQ ID NO 8  
; LENGTH: 9010  
; TYPE: DNA  
; ORGANISM: Human immunodeficiency virus type 1  
; FEATURE:  
; OTHER INFORMATION: Isolates=96ZM651; 137..1621:"gag"; 1426..4425:"pol";  
; OTHER INFORMATION: 4370..4948:"vif"; 4888..5178:"vpr";  
; OTHER INFORMATION: 5159..5373-7734..7824:"tat"; 5298..5373-7734..7981:"rev";  
; OTHER INFORMATION: 5387..5647:"vpu"; 5565..8171:"env"; 8173..8793:"nef"  
US-09-184-418C-8

Query Match 48.9%; Score 1203.6; DB 4; Length 9010;  
Best Local Similarity 70.0%; Pred. No. 3.2e-193;  
Matches 1650; Conservative 0; Mismatches 699; Indels 7; Gaps 2;  
Qy 14 TGGCGGAGGCGCATGAGCGCCACAGCGCCCAACATCTCTGATGCGAGCGCAACTTCA 73  
Db 1221 TGGCTGAGGCAATGAGCCCAACAAATAGTGTAACTACTGATGACAGAAAGCAATTTA 1280  
Qy 74 AGGCGCCCAAGCGCATCATCAAGTGTCTCACTCGGCGAGGCGCCACATGCCCCGCA 133  
Db 1281 AAGGAAATAAAGAAATGTTAAATGTTTAACTGTGTAAAGAGGGGCACATAGCCAGAA 1340  
Qy 134 ACTGCGCGCGCCCGCGCAAGAGGCTGCTGGAAGTGGCGCAAGAGGCGGCCACCATGTA 193  
Db 1341 ATTGAGGCGCCCTAGGAAAAGGGCTGTGGAATGTGGAAGAGGAGGACACCAATGA 1400  
Qy 194 AGGACTGACCGAGCGCCAGGCGCAACTTCTTCCGAGGAGCTGGCTTCCCGGAGCA 253





RESULT 12  
 US-09-184-418C-9  
 ; Sequence 9, Application US/09184418C  
 ; Patent No. 6492110  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hahn, Beatrice  
 ; APPLICANT: Gao, Feng  
 ; APPLICANT: Shaw, George  
 ; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN  
 ; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1  
 ; FILE REFERENCE: D6287  
 ; CURRENT APPLICATION NUMBER: US/09/184,418C  
 ; CURRENT FILING DATE: 1999-11-02  
 ; NUMBER OF SEQ ID NOS: 112  
 ; SEQ ID NO 9  
 ; LENGTH: 8972  
 ; TYPE: DNA  
 ; ORGANISM: Human immunodeficiency virus type 1  
 ; FEATURE:  
 ; OTHER INFORMATION: isolate=96ZM751.3; 137.1632:gag; 1419.4435:pol;  
 ; OTHER INFORMATION: 4380.4958:vif; 4898.5188:vpr; 5169.7814:tat;  
 ; OTHER INFORMATION: 5308.7938:rev; 5407.5667:vpu; 5585.8128:env;  
 ; OTHER INFORMATION: 8130.8753:nef  
 US-09-184-418C-9  
  
 Query Match 47.6%; Score 1172.6; DB 4; Length 8972;  
 Best Local Similarity 69.7%; Pred. No. 4.9e-188;  
 Matches 1655; Conservative 0; Mismatches 694; Indels 26; Gaps 4;  
  
 QY 14 TGGCGAGGCGCATGAGCAGGCGCCAGCGCCACATCTCTGATCGAGCGAGCAACTTCA 73  
 DB 1214 TGGCTGAGCAATGAGCGCAAGTAAACATAACAACATAATGATGCAGAAAAGCAATTTA 1273  
 QY 74 AGGGCCCCAGGGCATCATCAAGTCTCACTGCGGAGGAGGCGGCACATCGCCGCA 133  
 DB 1274 AGGGCCCCAAAGAAATGTTAAATGTTTCACTGTGCGGAGGAGGCGGCATATAGCCAGGA 1333  
 QY 134 ACTGCGGCGCCCCCGCGCAAGAGGCGTCTGGAAGTGGCGAAGGAGGCGGCACACATGA 193  
 DB 1334 ATTGAGGCGTCTCTGGGAAAAAGGCTGTGGAATGTGGAAGGAGGACACCAATGA 1393  
 QY 194 AGGATGACCGAGCGCGAGCGCACTTCTCCGAGGAGGCGTCTCTCTCTCTCTCTCTCTCT 253  
 DB 1394 AAGATGCTACTGAGAGCAGGCTAA-TTTTATGGGAAAAATTTGGGCTTCTCTCTCTCTCTCT 1452  
 QY 254 AGGCGCGGAGTTCCCGAGGCGAGCAG-----AACCGCGCCCAACAGCC 295  
 DB 1453 AGGCGCGGGAACCTTCTTCAGAACAGACGAGCCACAGCCCGCCAGCTCCCAACAGCC 1512  
 QY 296 CCACAGCGCGAGCTGCGAGTGGCGCGGACACCCCGCGAGCGGCGGCGCGGCGGCGGCGGCG 355  
 DB 1513 CCACAGCAGAGAGCTTCAGGTTCCAGGAGCAACCCCTGCCCCGAGCGAGGAGGAGGAGAA 1572  
 QY 356 GCCAGGCGCACCTGAACTTCCCGCAGATCACCTGTGCGAGCGCCCGCTCTGTGAGCATCA 415  
 DB 1573 GACAAAGGAACCTTAACTGCGCTCAATCACTCTTTGGCAGCGACCCCTTGTCTCAATAA 1632  
 QY 416 AGGTGGGCGCGAGATCAGAGGCGCTCTGGAACACCGGCGCGAGCAGACCGCTGCTGG 475  
 DB 1633 AAGTAGGGGTGAGATAAGAGGCTCTCTTGGATACAGGAGCAGATGATACAGTATTAG 1692  
 QY 476 AGGAGATGAGCTGCGCGCAAGTGGAGCCAGCCAGATGATCGCGGCATCGCGGCTTCA 535  
 DB 1693 AAGAAATAAATTTGGCAGGAAAAATGMAAACCAAAATGATAGGAGGAATTTGAGGTTTAA 1752  
 QY 536 TCAGGTGCGCGAGTACGACAGATCTCTGATCGAGATCTGCGGAGGAGGAGGAGGAGGAGGAG 595  
 DB 1753 TCAGGTGAGACAGTATGATCAATCTATATAGAAATTTTGGGAAAAAGGCTATAGGTA 1812  
 QY 596 CGGTGCTGATCGGCG 655  
 DB 1813 CAGTATTAGTAGGACCTACCTGTCAACATAAATTTGGGAGAAATATGTTGACCGAGCTTG 1872

QY 656 GCTGCAACCTGAACTTCCCATCAGCCCATCGAGAGCCGTGCGCCGTGAAGCTGAAGCCCG 715  
 DB 1873 GCTGCACACTAAATTTTCCAAATAGTCTTATGAAACTGTACAGTAAATTAAGACCCAG 1932  
 QY 716 GCATGACGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGAGCCCTGA 775  
 DB 1933 GAATGGATGCCCAAGGGTCAAAACATGGCCATTGACAGAGAAAAATAAAGCATTA 1992  
 QY 776 CCGCATCTCGGAGGAGTGGAGAGGCGCAAGATCACCAAGATCCCAAGATCGCCCGGAGACC 835  
 DB 1993 CAGCAATTTGTGAAGAAATGGAAGAGGAGGAAAAATTAACAAATTTGGCCCTGAGATC 2052  
 QY 836 CCTACAAACACCCCGCTGTGCGCATCAAGAAGAGAGCAGCACCAGTGGCGCAAGCTGG 895  
 DB 2053 CATATAACACTCCAGTATTTGCCATAAAAAAGAGCAGTACTAAGTGGAGAAAAATTAG 2112  
 QY 896 TGGACTTCCGCGAGCTGAACAGCGCACCCAGGACTCTTGGGAGGTGCGAGCTGGCATCC 955  
 DB 2113 TAGATTTTCAGGGAACCTCAATAAAGAACTCAGAGCTTTTGGGAAGTCAATTAGGAATAC 2172  
 QY 956 CCCACCCCGCGCGCTGGAAGAGAGAGAGCGTGTACCGTGTGACGCGGCGCAGCGCT 1015  
 DB 2173 CACACCCAGCGGGTTAAAAAAGAAAAAGTCACTGATGTGGGGGATGGT 2232  
 QY 1016 ACTTCAGCGTGGCGGAGGAGTTCGGAAGTACACCGCTTCACCATCCCGCAGCA 1075  
 DB 2233 ATTTTCAGTTCCTTAGATGAAGGCTTCAGGAATATATCTGATTCACATACCTAGTA 2292  
 QY 1076 TCAACACAGAGACCCCGGCATCCGTACAGTACAACTGTCTGCCCGCAGGCTGGAAG 1135  
 DB 2293 TAAACAATGAAACACCTGGGATTAGATATCAATATATATGCTTCCACAGGAGTGAAG 2352  
 QY 1136 GCACCCCGCAGTCTTCCAGAGCAGATGACCAAGATCTGTGAGCGCTTCCCGCGCGCA 1195  
 DB 2353 GATCACATCATATTCAGAGTAGCATATAAATCTTAGAGCCCTTTAGAGACAAA 2412  
 QY 1196 ACCCGAGATCGTGTATACCA-----GGCCCCCTGTACGTGGCGAGCGACCTGGAGA 1249  
 DB 2413 ACCCAGAAATAGTTATCTATCAATATATGATGATGATGATGATGATGATGATGATGAT 2472  
 QY 1250 TCGCGCAGCAGCCCGCAGATCGAGAGCTGCGCAGCAGCTGCTGCGCTCGCGCTTCA 1309  
 DB 2473 TAGGGCAACACAGACMAAANTAGAGGTTAAGAACACCTATTGAGATGGGANTTA 2532  
 QY 1310 CCACCCCGCAAGAAGCACAGAGAGCGCCCTTCTGTGTGATGGGCTACGAGCTGC 1369  
 DB 2533 CTACACAGCAAGAGCATCAGAAAGAGCGCCCAATTTCTTTGGATGGGCTATGAATCC 2592  
 QY 1370 ACCCGAGAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGCGCGTGA 1429  
 DB 2593 ATCTGACAAATGACACGATACAGCTTAAAGCTGCCAGAAAGAGAGCTGAGCTGTCA 2652  
 QY 1430 ACAGCATCCAGAGCTGTGGGCAAGCTGAACCTGGGCGCAGCAGATCTAACCCCGCATCA 1489  
 DB 2653 ATGATATACAGAAAGTTAGTGGGAAAAATTAACCT-GGCAAGTCAGATTTACGAGGGAITTA 2711  
 QY 1490 AGGTGCCCGAGCTGTGCAAGCTGCTGCGGGCGCAGGCGCTGACCGACATCTGTGCCCC 1549  
 DB 2712 AAGTAAGCACTGTGTAACTCTTAGGGAGCCAAAGCATTAACAGACATAGTACCAT 2771  
 QY 1550 TGACCGAGGAGCGAGCTGGAGCTGGCCAGAACCCGAGATCTCTGCGCGAGCGCGCTGC 1609  
 DB 2772 TGACTGAAGAGGCGAGAAATTAAGATTTGGCAGAGCAGGAGGAAATTTCTAAAGAACCCAGTAC 2831  
 QY 1610 ACAGGCTGTACTACGACCCCGCAGCAGGAGCTGGTGGCGAGATCCAGAGAGCGGCGCAGC 1669  
 DB 2832 ATGAGTATATTATGACCCATCAAGAGCTTAATAGTGAATACAGAAACAGGCGATG 2891  
 QY 1670 ACCAGTGAACCTACAGATCTACAGAGCGCTTCAAGAACCTTCAAGAGCGCGCAGTACG 1729  
 DB 2892 ACCAATGGACATATCAAGTTTACCAAGAACCAATTTCAAAATCTGAAACACAGGAAATG 2951

QY 1730 CCAAGTGGCCGACCGCCACCAACGACGTGAAGCAGCTGACCGAGCCGCTGCAGAGA 1789  
Db 2952 CAAATATGAGGAGCTGCCACACTAATGATGTAACAGCTTACAGAGCGGTGCAGAAAA 3011  
QY 1790 TCOCCTGGAGAGCATCGTGAATCGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGA 1849  
Db 3012 TAGCCATGGAAGCATAGTAATATGCGGAAGATTCTAAATTTAGGCTACCCATTCAA 3071  
QY 1850 AGAGACCTGGGAGACCTGGTGAACGACTACTGCGAGCGCCACCTGGATCCCGAGTGG 1909  
Db 3072 AAGAACAATGGGAGACATGTTGACAGACTATTTGGCAAGCCACCTGGATTTCTGAGTGG 3131  
QY 1910 AGTTCTGTAAACACCCCCCTGGTGAAGCTGTGTGTACAGCTGGAGAGAGGCCATCA 1969  
Db 3132 AGTTTGTTAATCTCCCTCCCTAGTAAATTTATGTTACAGCTGGAGAAAGAACCCATAG 3191  
QY 1970 TCGCGCCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGCAAG 2029  
Db 3192 CAGGAGCAGAACTTACTATGTAGATGGAGCAGCCCAATAGGGAACTAAATAGGAAAG 3251  
QY 2030 CCGGCTACGTGACCGACCGGCGCGGCGGAGAGATCGTGAGCCTGACCGAGACCAAC 2089  
Db 3252 CAGGCTATGTACTGACAGAGGAAGCAAAAAATTTGTTACTTAATGAAACAAATC 3311  
QY 2090 AGAAGACCGAGCTGCGAGCGCCATCCAGCTGGCCCTGCGGACAGCGCGAGCGGTGACA 2149  
Db 3312 ARAAGACTGAATACAGCAATTCAGTTAGCTTTGAGGATTCAGATCAGAGTAACA 3371  
QY 2150 TCGTGACCGACCGCAGTACGCTGGGCTATCATCCAGCCCGAGCGCGAGCGAGTGA 2209  
Db 3372 TAGTAACAGACTCACAGTATGCAATAGGAATCATCAAGCAACACAGATGAAT 3431  
QY 2210 GCAGCTGTGTGAACAGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTACCTGAGCT 2269  
Db 3432 CAGAAATGATCAATCAATTAATAGAACAGTTGATTAAGAAAGGAGTTTACCTGTAT 3491  
QY 2270 GGGTGGCCCGCCACAGGCGATCGGCGCAACGAGCAGATCAGAAAGCTGTGTAGCAAG 2329  
Db 3492 GGGTACGACACACAAAGAAITGGAGGAAATGAACAAGTAGATAAATGGTAAAGTAG 3551  
QY 2330 GCATCGCAGAGTGTCTCTCGAGCGCATCGAT 2364  
Db 3552 GAATCAGGAAGTGTCTTCTAGATGGAATAGAT 3586

RESULT 13  
US-09-184-418C-11  
; Sequence 11, Application US/09184418C  
; Patent No. 6492110  
; GENERAL INFORMATION:  
; APPLICANT: Hahn, Beatrice  
; APPLICANT: Gao, Feng  
; APPLICANT: Shaw, George  
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN  
; FILE REFERENCE: D6287  
; CURRENT APPLICATION NUMBER: US/09/184,418C  
; CURRENT FILING DATE: 1999-11-02  
; NUMBER OF SEQ ID NOS: 112  
; SEQ ID NO 11  
; LENGTH: 8959  
; TYPE: DNA  
; ORGANISM: Human immunodeficiency virus type 1  
; FEATURE:  
; OTHER INFORMATION: isolate=94IN476.104; 138.1613: "gag";  
; OTHER INFORMATION: 1418.4428: "pol"; 4361.4939: "vif"; 4879.5169: "vpr";  
; OTHER INFORMATION: 5150.7782: "tat"; 5289.7939: "rev"; 5378.5638: "vpu";  
; OTHER INFORMATION: 5556.8129: "env"; 8131.8754: "nef"  
US-09-184-418C-11

Query Match 47.3%; Score 1165.4; DB 4; Length 8959;  
Best Local Similarity 69.6%; Pred No. 7, 9e-187;  
Matches 1640; Conservative 0; Mismatches 706; Indels 11; Gaps 4;

QY 14 TGGCGAGGCCATGAGCCAGCCACCGAGCCCAACATCTCTGATCGAGCGCAACTTCA 73  
Db 1216 TGGCTGAGGCAATGAGCCAAATCACATAG---TAAACATAAATGATGACAGAGGCAATTTTA 1272  
QY 74 AGGGCCCCAAGCGCATCATCAAGTGTCTTCACTCGGCAAGGAGGCGCAATCGCCCGCA 133  
Db 1273 AAGGCCCTTAAAGAAATTTGTTAAATGCTTCAATCTGTGCAAGGAAGGCAATAGCCAGAA 1332  
QY 134 ACTGCGCGCCCCCGCGAAGAGGCTGCTGGAAGTGGGCAAGAGGGGCCACACAGATGA 193  
Db 1333 ATTGCAAGGCCCTTAGAAAAAGAGGCTGTGGAATTTGGGCAAGAAAGGACACCAATGA 1392  
QY 194 AGGACTGCAACGAGCGCGCAGGCCAACTTCTCCGCGAGGACCTGGCTTCCCGCAGGCA 253  
Db 1393 AAGACTGTACTGAGAGGCGAGGCTTAA-TTTTTAGGAAATTTGGCTTCCCAAGGGG 1451  
QY 254 AGCGCGCGAGTTCCCGAGCGAGCAACCGCGCAACAGCCCCCAACAGCGCGCGAGTGC 313  
Db 1452 AGCGCAGGGAATTTCTTCAAAAAGGCGAGGCAACAGCCCCCAGCAGAGAGGCTTC 1511  
QY 314 AGGTGCGGGGCAACACCCCGCAGCGCGCGCGCGCGAGCGCGCGAGCGCGCACTGAACT 373  
Db 1512 AGGTTCAGGAGCAACACCCCGCTCCGAGCAGAGTCTGAAGACAGGGAACCTTAACT 1571  
QY 374 TCCCGCAGATACCTGTGGCAGCGCCCTGTGTGAGCATCAAGTGGCGCGCGCAGATCA 433  
Db 1572 TCCCTCAATCACTCTTTGGCAGCGACCCCTTGTCTCAATAAAAGTAGGGGCGCAGATA 1631  
QY 434 AGAGCGCTCTGTGACACCGCGCGCGAGCAACCGTGTGGAGAGATGAGCCTGCGCG 493  
Db 1632 AGGAGCTCTCTGACACAGGAGCAGATGATACAGTATTAGAAATAGCTTTGCGAG 1691  
QY 494 GCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAGTGGCGCGCAGTAC 553  
Db 1692 GAAGTGGAAACCAAAAATGATAGAGGAATTTGAGGTTTTTATCAAAAGTAGACAGATG 1751  
QY 554 ACCAGATCTGATCGAGATCTCGCGCAAGAGGCGCATCGGCGCGCTGCTGATCGCGCCCA 613  
Db 1752 ATCAATATCTTAGAATTTGTGAAAAGGCTATAGGTACAGTATTAGTAGGACCTA 1811  
QY 614 CCCCCTGAAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGACCCCTGAACTTCC 673  
Db 1812 CACCTGTCAACATTAATTGGAAGAGATATGTTGACTCAGCTTGGATGCACTCTAAATTTTC 1871  
QY 674 CCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCGCGCGCATCGAGCGCCCAAGG 733  
Db 1872 CAATTAGCCCCATTTGAACCTGTACAGTAAATTTAAAGCCAGGAATGATGGGCCCAAGG 1931  
QY 734 TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCCATCTGGGAGAGA 793  
Db 1932 TTAACAGTGGCCATTGACAGAGAGAAATATAAGCATTAACAGAAATTTGTAAGAAA 1991  
QY 794 TGGAGAAGGGGCAAGATCACCAAGATCGCGCGCGAGACCCCTACACACACCCCGTGT 853  
Db 1992 TGGAGAAGGAAGAAATTAACAAAATTTGGGCTGAAAATTCATATTAACACTCCAGTAT 2051  
QY 854 TCGCCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913  
Db 2052 TTGCCATAAAAGGAAGGACAGTACTAAGTGGAGAAAATTTAGTAGATTTTCAGGAGCTCA 2111  
QY 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCACTCCCGCCCGCGGCTGA 973  
Db 2112 ATAAAGAACTCAAGACTTTTGGGAAGTTCAATTAGGAATACCAACCCAGAGGTTAA 2171  
QY 974 AAGAAGAAGAGCGTGAACCGTGTGACGCGTGGGCGAGCGCTACTTTCAGCGTCCCGCTGG 1033  
Db 2172 AAAAGAAAAATCAGTGACAGTCTGATGTGGGGATGCAATATTTTCAGTTCCCTTTAG 2231  
QY 1034 ACGAGGACTTCGCAAGTACACCGCTTACCACTCCCGCAGCATCAACAGAGACCCCGG 1093  
Db 2232 ATGAAGGCTTCGGGAATATTAATCTGATTCACCATACCTAGTATATAACATGAACACCG 2291

1094 GCATCGCTACAGTACCAAGCTGTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153  
Db  
2292 GGTATAGATATCAATATATATGCTTCCACAGGGATGGAAGGATCACCAGCATATTC 2351  
QY 1154 AGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGGCCCGCCGCAACCCCGAGATCGTGATCT 1213  
Db 2352 AGAGTAGCATGACCAAAATCTTAGAGCCCTTTAGAGCCGCAAGAAATCCAAAATATAGTCATCT 2411  
QY 1214 ACCA-----GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCAGCACCCCGGCCA 1267  
Db 2412 ATCAATATATGATGACTGTATGATGAGGTCTGACTTAGAATAGGCATCATAGACAA 2471  
QY 1268 AGATCGAGAGCTGCGCAAGCACTGTCTGCGCTGGGGCTTACCAACCCCGCAGCAGAGC 1327  
Db 2472 AAATAGAGGAGTTAAGAGACATCTATTAAAGTGGGGATTACCAACACCAAGATAGAAAC 2531  
QY 1328 ACCAAGAGAGCCCCCTCTCTGTGGATGGGTGCTAGAGCTGACCCCGCCGCAAGTGGACCG 1387  
Db 2532 ATCAGAAAGAACCCCAATCTCTTGATGGGTATGAATCCATCTCTGACAAATGGACAG 2591  
QY 1388 TCAGCCCATCAGCTGCCGAGAGAGAGCTGGACCGTGAAGCAGATCCAGAGCTGG 1447  
Db 2592 TAGAGCTTAAGCTGCCAGAAAGGATAGTGGACTGTCAATGATATACGAAGTTAG 2651  
QY 1448 TGGCAAGCTGAACCTGGGCAGCCAGATCTACCCCGCATCAAGGTGGCCAGCTGTGCA 1507  
Db 2652 TGGGAAATTAACCTGGGCAAGTCAAGTTTACCAGGGATTAAAGTGAAGCACTTTGTA 2711  
QY 1508 AGCTGCTGGCGGCGCCAGGSCCTTGACGACATGCTGCCCCCTACCGAGAGGCCGAGC 1567  
Db 2712 AGCTCTTTAGGGGGCCAAAGCACTTACAGACATAGTACCCTAACTAACTGAAGAGCAGAA 2771  
QY 1568 TGGAGTGGCCGAGAACCCGAGATCTCTGCGGAGCCGCTGACGGGTGTACTACGACC 1627  
Db 2772 TAGAATTAGCAGAGAACGGGAAATCTTAAAGAGCCAGTACATGGAGTATATTATGACC 2831  
QY 1628 CCAGCAGAGCACTGTGGCCGAGATCCAGAGCAGGGCCAGCAGCTGACCTTACCAGA 1687  
Db 2832 CATCAAGACACTTAATAGCTGAAATACAGAAACAGGGGCATGACCAATGGACATATCAA 2891  
QY 1688 TCTACAGAGGCGCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAGATGCGCACCCGCC 1747  
Db 2892 TTACCAAGAACCAATCAAAATCTGAAAACAGGGAAGTATGCAAAATGAGGACTGCTC 2951  
QY 1748 ACACCAACAGCTGAGCAGCTGACCGAGGCGCTGCGAAGATGCCATGAGAGCATCG 1807  
Db 2952 ACCTAATGATGATAAAGCTTAAAGAGGCGAGTGCAGAAATAGCCATAGAAAGCATAG 3011  
QY 1808 TGATCTGGGCGAAGACCCCAAGTTCCGCTGCCATCCAGAGGAGACCTGGGAGACCT 1867  
Db 3012 TAATAT-GGGAAGACCCCTAAATTTAGACTACCCATCCAAAGAAACGTTGGGAGACAT 3070  
QY 1868 GGTGACCCGACTACTGGCAGGCCACTGGATCCCGAGTGGAGTTGCGNAACCCCCC 1927  
Db 3071 GGTGGACAGACTATTGGCAGGCCACTGGATCTCTGATTGGGAGTTTGTATACCCCTC 3130  
QY 1928 CCTGTGTGAAGCTGTGGTACCAAGCTGGGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1987  
Db 3131 CCTAGTAAATATTGGTACCAGCTAGAAAAGAAACCCATAGTAGGAGCAGAACTTTCT 3190  
QY 1988 AGTGGAGCGCGCCGACACCGCAGACCAAGATCGGCAAGCGCGCTAGTGACCGAAC 2047  
Db 3191 ATGTAGATGAGCAGCTAATAGGAACTAATAGTAGAAAGCAGGGTATGTTACTGACA 3250  
QY 2048 GGGGCGCGCAGAAAGTCTGTAGCCCTGACCGAGACCAACCAAGAGACCGAGCTGAGG 2107  
Db 3251 GAGGAAGGCAGAAAATTGTTTCTTTAACTGAAAACAAATCAGAAAGTGAATTCGAAAG 3310  
QY 2108 CATCCAGCTGCGCTGACAGGACAGCGCAGGAGTGAACATCGTACCGCAGCAGCCAGT 2167  
Db 3311 CAATTCAGCTAGCTTTGCAAGATTGAGAAACAGAAATGAATATAGTACAGCTACAGT 3370  
QY 2168 AGCCCTTGGGCATCATCCAGGCCCCAGCCGACAAAGAGCGAGCGAGCTGGTGAACCGA 2227

3371 ATGATTAGGATCATTTCAAGCAACACGATAAAGTGAATCAGAGTTAGTCAACCAA 3430  
QY 2228 TCATCGAGCAGCTGATCAAGAGGAGAGTGTACTCTAGCTGGGTGCGCCGCCACAGG 2287  
Db 3431 TAATAGAACATTAATAAACAAGAAAGATCTTATCTGTCTGGGTACCAACACATAAG 3490  
QY 2288 GCATCGGCGGCAACGACGACATCGAACAGCTGTGAGCAAGGCGCATCCCGCAAGGTGCTGT 2347  
Db 3491 GAAATTGGAGGAATGAACAAGTAGATAGATTAGTAAGTAGTGAAATTAGGAAAGTACTGT 3550  
QY 2348 TCTTGGAGCGCATCGAT 2364  
Db 3551 TTCTAGATGGGATAGAT 3567

## RESULT 14

US-09-872-733A-3  
; Sequence 3, Application US/09872733A  
; Patent No. 6656706  
; GENERAL INFORMATION:  
; APPLICANT: The Government of the United States of America, as  
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND  
; FILE OF INVENTION: SIV ENV GENES  
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL, SIV GAG & ENV  
; CURRENT APPLICATION NUMBER: US/09/872,733A  
; CURRENT FILING DATE: 2001-06-01  
; PRIOR APPLICATION NUMBER: PCT/US00/34985  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/173,036  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 2467  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURES:  
; OTHER INFORMATION: Description of Artificial Sequence: Mutated Human  
; OTHER INFORMATION: Immunodeficiency Virus - 1 Pol gene  
US-09-872-733A-3

Query Match 47.3%; Score 1163.8; DB 4; Length 2467;  
Best Local Similarity 84.0%; Pred. No. 1.2e-186;  
Matches 1328; Conservative 0; Mismatches 247; Indels 6; Gaps 1;  
QY 790 GAGATGGAGAAGGAGGCGGCAAGATCAACAGATCGGCCCGAGAACCCCTACACACTCCA 849  
Db 7 GAGATGGAGAAGGAGGAGGAGATCAGCAAGATCGGCCCTGAGAACCCCTACACACTCCA 66  
QY 850 GTGTTTGGCCATCAAGAAGAGGAGCAGCAGCAAGTGGCGCAAGCTGGTGGACTTCCCGCAG 909  
Db 67 GTCTTCGCATCAAGAGAGAGGAGCAGTACCAGTGGAGAAAGCTGTGGACTTCAGAGAG 126  
QY 910 CTGAACAAAGCCACCCAGGACTTCTGGAGGTGAGCTGGGGATGCCCCACCCCGCCGCG 969  
Db 127 CTGAACAAAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATCCCATCCCGCTGGG 186  
QY 970 CTGAAGAAGAAAGAGAGCGTGACCGTCTGAGCGTGGGGCAGCGCTACTTTCAGCGTGC 1029  
Db 187 TTGAAGAGAGAAAGTCACTGAGTGTGGTGTGATGCTCTACTTCTCCGTGCC 246  
QY 1030 CTGAGCAGGAGCTTCGGAGAGTACACCGCTTTCACCATCCCGAGCATCAACACGAGACC 1089  
Db 247 TTGAGCAGGAGCTTCAGGAAAGTACCTGCGCTTCCAGTACCTAGCATCAACACGAGACA 306  
QY 1090 CCGGCATCCCTACCACTACAGTCAACGCTCTGCCAGGGCTGGGAAGGCGACGCCCGCATC 1149  
Db 307 CCAGGATCCCGTACCACTACAGTCAACGCTCTGCCAGGGATGGAAGGATCACCAGCCATC 366  
QY 1150 TTCAGAGCAGCATGACCAAGATCTCTGAGCGCTTCCGCGCCCGCAACCCCGAGATCGTG 1209  
Db 367 TTTCAGAGCAGCATGACCAAGATCTCTGAGCGCTTCCGCAAGCAAAACCCAGCATCTGTG 426

QY 1210 ATCTACCA-----GSCCCCTGTACGTGGGCGAGCGCTGGAGTGGCCGACGACCGC 1263  
Db 427 ATCTATCAGTACATGAGCACTCTACGTAGAGAGTGACCTGGAGATCGGCGACGACAGG 486  
QY 1264 GCCAAGATCGAGAGTGGCGAAGCACTCTGCGCTGGGGCTTCAACACCCCGCAAG 1323  
Db 487 ACCAAGATCGAGAGTGGCGAAGCACTCTGCGCTGGGGCTTCAACACCCCGCAAG 546  
QY 1324 AAGCACGAGAAGAGCGCCCTTCTGTGTGGTGGCTACGAGCTGCACCCCGCAAGTGG 1383  
Db 547 AAGCACGAGAAGAGCGCCCTTCTGTGTGGTGGCTACGAGCTGCACCCCGCAAGTGG 606  
QY 1384 ACCGTGAGCCCATCGAGTGGCGGAGAGAGAGTGCACGTAAGCAATCCAGAG 1443  
Db 607 ACAGTCAGCCCATCGTGTGCTGAGAGAGAGAGTGCACGTAAGCAATCCAGAG 666  
QY 1444 CTGGTGGCAAGCTGAATGGCCAGCAGATCTACCCCGGATCAAGGTGGCGCAGCTG 1503  
Db 667 CTGCTGGCAAGTGGTGAATGGCCAGCAGATCTACCCCGGATCAAGGTGGCGCAGCTG 726  
QY 1504 TGAAGCTGTGGCGGCGCCAGAGCCCTGACGACATCTGTGCCCTGACCGAGAGGCC 1563  
Db 727 TGAAGCTGTGGCGGCGCCAGAGCCCTGACGACATCTGTGCCCTGACCGAGAGGCC 786  
QY 1564 GAGCTGAGTGGCCGAGAACCGGAGATCTCTGCGGAGCCGCTGCACGGGCTGTACTAC 1623  
Db 787 GAGCTGAGTGGCCGAGAACCGGAGATCTCTGCGGAGCCGCTGCACGGGCTGTACTAC 846  
QY 1624 GACCCAGAGAGACCTGTGGCGGAGATCCAGAGAGAGGCGACGACATGTGACCTAC 1683  
Db 847 GACCCAGAGAGACCTGTGGCGGAGATCCAGAGAGAGGCGACGACATGTGACCTAC 906  
QY 1684 CAGATCTACAGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACCGCAAGATCGCAC 1743  
Db 907 CAAATCTACAGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACCGCAAGATCGCAC 966  
QY 1744 GCCCAACCAACGAGTGAAGCAGTGCACGAGGCGGTGCAGAGATCGCCATGAGAGC 1803  
Db 967 GCCCAACCAACGAGTGAAGCAGTGCACGAGGCGGTGCAGAGATCGCCATGAGAGC 1026  
QY 1804 ATGCTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGAGACCTGGGAG 1863  
Db 1027 ATGCTGATCTGGGGCAAGACCTCCCAAGTTCAAGTGCCTATACAGAGAGATGGAG 1086  
QY 1864 ACTGTGACCGACTATCTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCGTGAACAC 1923  
Db 1087 ACATGTGACCGACTATCTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCGTGAACAC 1146  
QY 1924 CCCCCCTGGTGAAGCTGTGTACAGCTGAGAGAGAGCCCATCATCGCGCGCGAGACC 1983  
Db 1147 CTTCCCTTGGTGAAGCTGTGTGTATCAGCTGAGAGAGAGCCCATCATCGTGGAGCAGACC 1206  
QY 1984 TTCTACGTGAGCGGCGCCCAACCGGAGACCAAGATCGGCAAGCGCGGCTACGTGACC 2043  
Db 1207 TTCTACGTGAGTGGGCGACCAACAGGAGACCAAGCTGGGCAAGCGAGCTACGTGACC 1266  
QY 2044 GACCGGCGCGGAGAGATGTGAGCTGACCGAGACCAACCAAGAGAGCGAGCTG 2103  
Db 1267 AACCGAGGAGCAGAGAAAGTGTGACCTGATGACACCAACCAAGAGAGCTGAGCTG 1326  
QY 2104 CAGGCCATCAGCTGCGCTGAGGACAGCGGAGAGAGTGAACATCGTGACCGAGCAGC 2163  
Db 1327 CAAGCCATCTACCTAGCTGCAAGACAGCGGAGTGAAGTGAACATCGTGACAGACTCA 1386  
QY 2164 CAGTACGCTGGGATCATCCAGGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2223  
Db 1387 CAGTACGCTGGGATCATCCAGGACCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1446  
QY 2224 CAGATCATCGAGCAGCTGATCAAGAGAGAGAGAGTGTATCCTGAGCTGGGTGCCCGCAC 2283  
Db 1447 CAGATCATCGAGCAGCTGATCAAGAGAGAGAGAGTGTATCCTGGCATGTGTTACCGACAC 1506

QY 2284 AAGGCATCGCGGCAACGACGATCGACAAGCTGGTGGCAAGGCGCATCCGCAAGGTG 2343  
Db 1507 AAGGAATGGAGGAATGAACAGTAGATTAATTAATCTAGTCTGGATCCGAGAGTG 1566  
QY 2344 CTGTTCTCGGAGCGCATCGAT 2364  
Db 1567 CTGTTCTCGGAGCGGATCGAT 1587  
RESULT 15  
US-09-184-418C-4  
; Sequence 4, Application US/09184418C  
; Patent No. 6492110  
; GENERAL INFORMATION:  
; APPLICANT: Hahn, Beatrice  
; APPLICANT: Gao, Feng  
; APPLICANT: Shaw, George  
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN  
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1  
; FILE REFERENCE: D6287  
; CURRENT APPLICATION NUMBER: US/09/184,418C  
; NUMBER OF SEQ ID NOS: 112  
; SEQ ID NO 4  
; LENGTH: 8992  
; TYPE: DNA  
; ORGANISM: Human immunodeficiency virus type 1  
; FEATURE:  
; OTHER INFORMATION: isolates=92RW009; 139.1624:989; 1690.4428:pol(N-terminus uncertain);  
; OTHER INFORMATION: 4373.4951:vif; 4891.5181:vpf; 5162.7801:cat; 5301.7958:rev;  
; OTHER INFORMATION: 5403.5648:vpu; 5566.8148:env; 8150.8773:nef  
US-09-184-418C-4  
Query Match 46.4%; Score 1142; DB 4; Length 8992;  
Best Local Similarity 68.6%; Pred. No. 6.5e-183;  
Matches 1620; Conservative 0; Mismatches 730; Indels 10; Gaps 3;  
QY 14 TGGCGAGGCGCATGAGCGGCCACCGAGCCCAACATCTCTGATGCGGAGGCACTTCA 73  
Db 1221 TGGCTGAAGCAATGAGCGCAAGTACAAACCAACATATATGATGAGAGGCAATTTTA 1280  
QY 74 AGGCGCCCAAGCGCATCATCAAGTGTCTCAACTCGGCAAGGAGGCGCATCGCCGCA 133  
Db 1281 AGGCGCAGAGAGAAATTAATTAAGTGTCTCACTGTGCAAGAGAGGACACTAGCAGAA 1340  
QY 134 ACTGCGCGCCCGCGCAGAGAGGCGCTGTGAAGTGGCGAAGAGGCGCCACAGATGA 193  
Db 1341 ATTGCGAGGCGCCCTAGAAAAAAGGGCTGTGGAATGTGGAAGAGGAGGACACCAATGA 1400  
QY 194 AGGACTGACCGAGCGCGCGAGCGCAACTCTTCCCGAGGACCTGGCCTTCCCGAGGCA 253  
Db 1401 AAGACTGCACTGAGAGACAGGCTAA-TTTTATTAGGAAAAATTTTGGCCTTCCAAAGGGG 1459  
QY 254 AGGCGCGGAGTTCCCGAGCGAGAGAACCGCGCCAAACAGCCCGCACAGCGCGAGCT-- 311  
Db 1460 AGGCGCAGGAAATTTTCCCGAGCGAGACTGAGAGCAACAGCCCGCACAGCAGAGAACTTT 1519  
QY 312 -CGAGGTGGCGGAGCAACCCCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 370  
Db 1520 GGAATGGGGGAAGAGATAGCCTCTCTCTGAAACAGGAGCAGAAAGACAGGAGACCTTTA 1579  
QY 371 ACTTCCCGCAGATCACCTGTGGAGCGCGCCCTCTGTGAGCATCAAGGTGGCGCGCAG 430  
Db 1580 ATTTCCCTCAAATCACTCTTTGGCAACGAGCCCTTGTCAAGTAAAAATAGGAGGTGAGC 1639  
QY 431 TCAAGAGGCGCTGTGAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 490  
Db 1640 TGAAGAGAGCTCTATTAGTACAGAGCAGATGATAGATTTAGAGAGAAATTAATTTGC 1699  
QY 491 CGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCGCAGT 550  
Db 1700 CAGGAAATGGAACCAAAAAATGATAGGGGAAATGAGGGGAAATTTGAGGTTTATCAAGGTAAAAACAGT 1759

QY 551 ACGACAGATCTGATCGAGATCTGGCGAAGAGGCGCATCGCCGCTGTGATCGGCC 610  
 Db 1760 ATGATCAAAATCTATAGAAAATTTGGGAAAGAGGCTATAGGTACAGTATTTAGTAGAC 1819  
 QY 611 CCACCCCGGTGAACATCATCTGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTGAAC 670  
 Db 1820 CTACATCTGTCAACATAATTTGGAAGAAATATGTTGACCCAGATGTTGTTGATCTTTAAACT 1879  
 QY 671 TCCCATACGCCCATCGAGACCGTCCGCTGAGCTGAGCTGAAGCCGCGCATGGACGCGCCCA 730  
 Db 1880 TTCCAAATAGTCTATTTAGAGCTGTACCAAGTAGCATTAAAGCAGGAATGGATGGCCCA 1939  
 QY 731 AGGTGAAGCAGTGGCCCTGTACCGAGGAGAGATCAAGGCCCTGTGACCGCCATCTCGGAGG 790  
 Db 1940 AGGTAAACATGCCCATTGACAGAGAAATAAAGCAATTAAGAGAAATTTGTACAG 1999  
 QY 791 AGATGGAAGAGGAGCAAGATCACCAAGATCGGCCCGAGACCCCTACACACCCCGC 850  
 Db 2000 AAATGGAAAGAGGGGAAATTTCAAATTCGGGCTCGAAATCCATATACATCCAG 2059  
 QY 851 TGTTCGCCATCAAGAGAAGAGCAGCACCAAGTGGCGCAAGCTGTGGACTTCCGCGAGC 910  
 Db 2060 TATTTGCCATAAAGAGAGGACAGTACTAAGTGGAGAAATTTAGTAGATTTCAAGGAAC 2119  
 QY 911 TGAACAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCTATCCGCCACCCCGCGGCC 970  
 Db 2120 TCAACAAAGAACTCAAGACTTTTGGGAAGTCCAAATTTAGGAGTACCACACCCAGAGGTT 2179  
 QY 971 TGAAGAGAGAGAGAGGCTGACCGTGTGACGCTGGGCGACGCTACTTTGAGCGTCCCC 1030  
 Db 2180 TAAAGAGAAATCACTGAGCAGTACTGGATGTGGGGATGCATCTTCTCAGTTCCCTT 2239  
 QY 1031 TGGACGAGGACTTCCGCAAGTACACCGCTTACCATCCCGGCGGCAACAGAGAGCC 1090  
 Db 2240 TAGATGAGGCTTCAGGAAATATATCTGATTCACATACCTAGTATAAACAATGAACAC 2299  
 QY 1091 CCGCATCCGCTACAGTACAACTGCTGCCAGGCTGGAGGGGAGCCCGCAGCATCT 1150  
 Db 2300 CAGAAATAGGTATCAATATATGTGCTTCCACAGGATGGAAGGATCACAGCAATAT 2359  
 QY 1151 TCCAGAGCAGCATACCAAGATCTCGAGGCTTCCGCGCGCGCAACCCGAGATGTGA 1210  
 Db 2360 TCCAAATAGTATGACAAAATCTTAGAGCCCTTTAGGGCACAACCAAGAAATAGTGA 2419  
 QY 1211 TCTACCA-----GGCCCCCTGTACGTGGGCGACGCTGGAGATCGGCGAGCACCGCG 1264  
 Db 2420 TCTATCAATATATGATGACTTTGACGTAGATCTGACTTAGAATAGGGCACTAGAG 2479  
 QY 1265 CCAAGATCGAGAGTGGCGAAGCACCTGTGCGCTGGGGCTTCAACCCCGGACAAAG 1324  
 Db 2480 CAAAAATAGAGGTTAAGAGAACATCTATTAAAGTGGGGATTTACCACACCCAGCAAGA 2539  
 QY 1325 AGCACCAGAGGAGGCCCCCTTCTGTGATGGGCTACGAGCTGCACCCGACAAAGTGA 1384  
 Db 2540 AACATCGAAGAAACCTCCATTTCTTTGGATGGGGTATGAACCTTCCTCTGACAAATGGA 2599  
 QY 1385 CCGTCAGCCCATCGAGCTGGCGGAGAGAGAGTGGACCGTGAACGACATCCAGAAC 1444  
 Db 2600 CAGTACACCTATACAGCTGCCAGAAAGGATAGCTGGACTGTCAATGATATACAGAGT 2659  
 QY 1445 TGGTGGGAGCTGAACTGGGCGACGAGATCTACCCGGCATCAAGTGGCGCCAGCTGT 1504  
 Db 2660 TAGTGGGAAATTTAAACTGGGCAAGTCAGATTTACCCAGGGGTTAAAGTAAGGCAATGT 2719  
 QY 1505 GCAAGCTGTGGCGGCGCAAGGCCCTGACCGACATCGTGGCCCTGACCCGAGGAGCGC 1564  
 Db 2720 GTAACTCTTAGGGGACCAAGCATTAACAGACATAGTACCATACTGAAGAGCAG 2779  
 QY 1565 AGCTGAGCTGGCGAGAACCCGAGATCTCGCGGAGCCCGTGCACCGCGTGTACTACG 1624  
 Db 2780 AATTAGAAATGGCAGAAACAGGAAATTTTAAAGAACCCAGTACATGGAGTATATATG 2839  
 QY 1625 ACCCCAGCAAGGACCTGTGGCGCGAGATCCAGAAAGAGGGCCACGACGAGTGGACCTACC 1684

Db 2840 ACCCATCAAAAGACTTTAATAGCTGAATACAGAAACAGGGGCTATGACCAATGGACATATC 2899  
 QY 1685 AGATCTACCAAGGAGCCCTTCAAGAACCTGAAGACCGCAAGTACGCCAAGATGCGCACCG 1744  
 Db 2900 AATTTTACCAAGAACCATTTCAAAAATCTGAAAACAGGAAGTATGCAAAAAGGAGGACTG 2959  
 QY 1745 CCCACAAACGAGCTGAAGCAGCTGACCGAGCCGCTGAGAGATCGCCATCGGAGAGCA 1804  
 Db 2960 CCCACATAATGACGTAAAAACGTTAACAGAGCAGTGCAGAAAGATAGCCATCGGAAGCA 3019  
 QY 1805 TCGTGTCTGGGCGACAGACCCCAAGTTCCGCTGCCCATCCAGAAAGAGAGACCTGGGAGA 1864  
 Db 3020 TAGTAAATATGGGAAAGACTTCTTAATTTAGATTTACCATCCAGAAAGAAACATGGGAAA 3079  
 QY 1865 CTTGGTGGACCGACTACTTGGCAGCCACCTGGATTCGCCAGTGGGAGTTCGTGAACACCC 1924  
 Db 3080 CATGGTGGACAGACTATTGGCAAGCCACCTGGATTCCTGAGTGGGAGTTTGTTAATACCC 3139  
 QY 1925 CCCCCTGGTGAAGCTGTGTGTACAGCTGGAGAGAGCCCATCATCGGCGCGGAGACCT 1984  
 Db 3140 CTCCTCTAGTAAATTAATTTATGTTACAGCTAGAGAAAGAACCCATATTAGGAGCAGAGACTT 3199  
 QY 1985 TCTACGTGGACGGCGCCCAACCGGAGACCAAGATCGGCAAGGCCGCTACGTGACCG 2044  
 Db 3200 TCTATGTAGATGGAGCAGCTAATCGGAAACTAAAATAGGAAAGCAGGATGTGTACTG 3259  
 QY 2045 ACCGGCGCGCGCAGAGATCGTGAAGCTGACCGAGACCAACCAAGAGAGAGGAGCTGC 2104  
 Db 3260 ACAGAGGAAGCAGAAATTTGTTCTTCTAAGTAAACCAACATCAGAGACTGAATTAAC 3319  
 QY 2105 AGGCCATCCAGCTGGCCCTGCGAGACAGCGGAGCGAGGTGAACATCGTGAACCGAGACC 2164  
 Db 3320 AAGCAATTCAGCTAGCTTTACAGGATTCAGGATCAGAAAGTAAACATAGTAACAGACTCAC 3379  
 QY 2165 AGTACGCCCTGGGCATCATCCAGGCCCGACAGAGAGAGGAGGAGGAGCTGGTGAACC 2224  
 Db 3380 AGTATGCAATAGGAATCATTCAGCAACACAGATAGCAGGATCGGAGGCACTCAATC 3439  
 QY 2225 AGATCATCGAGCAGCTGATCAAGAGAGAGAGGTGTACCTGAGCTGGGTGCCCCCACA 2284  
 Db 3440 AATAATAGAACAGTTAATAAAGGAAAGAGTCTACCTGTCATGGGTACCCAGCATA 3499  
 QY 2285 AGGCATCGGCGGCAACGAGCAGATCGACAGCTGGTGAAGCAAGGAGCTCCGCAAGGTGC 2344  
 Db 3500 AAGGAATTTGGAGGAATGAACAGATAGATAAATTAGTAAAGTGTAGTGAATCAGGAGGTGC 3559  
 QY 2345 TGTTCCTGACGGCATCGAT 2364  
 Db 3560 TGTTCCTAGATGGAATAGAT 3579

Search completed: April 10, 2004, 16:22:46  
 Job time : 127.333 secs





Db 121 CACATCGCCGCAACTGCGCGCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAG 180  
 QY 181 GGCACACAGATGAGGACTGACAGAGGCGCAGAGCCCAACTTCTTCGCGAGAGACCTGGCC 240  
 Db 181 GGCACACAGATGAGGACTGACAGAGGCGCAGAGCCCAACTTCTTCGCGAGAGACCTGGCC 240  
 QY 241 TTCCGCCAGAGGAGAGGCGCGAGTTCCCGAGAGAGAGAACCGCGCGCAACAGCCCAAC 300  
 Db 241 TTCCGCCAGAGGAGAGGCGCGAGTTCCCGAGAGAGAGAACCGCGCGCAACAGCCCAAC 300  
 QY 301 AGCCGCGAGCTGACAGTGGCGGCGCAACACCCCGAGAGAGGCGCGCGCGAGAGCCAG 360  
 Db 301 AGCCGCGAGCTGACAGTGGCGGCGCAACACCCCGAGAGAGGCGCGCGCGAGAGCCAG 360  
 QY 361 GGCACCTTGAATTTCCCGAGATCACTCTGTGGAGAGCGCCCTGTGTGAGCATCAAGGTG 420  
 Db 361 GGCACCTTGAATTTCCCGAGATCACTCTGTGGAGAGCGCCCTGTGTGAGCATCAAGGTG 420  
 QY 421 GCGCGCCAGATCAAGAGGCGCTGTGGACACCGCGCGCGAGAGAGCGCGCTGTGGAGGAG 480  
 Db 421 GCGCGCCAGATCAAGAGGCGCTGTGGACACCGCGCGCGAGAGAGCGCGCTGTGGAGGAG 480  
 QY 481 ATGAGCTTGGCGGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTTCATCAAG 540  
 Db 481 ATGAGCTTGGCGGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTTCATCAAG 540  
 QY 541 GTGCGCCAGTACGACAGATCTGATCGAGATCTGCGGCGAAGAGGCGATCGGCAACCGTG 600  
 Db 541 GTGCGCCAGTACGACAGATCTGATCGAGATCTGCGGCGAAGAGGCGATCGGCAACCGTG 600  
 QY 601 CTGATGCGGCGCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCGAGCTGGGCTGC 660  
 Db 601 CTGATGCGGCGCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCGAGCTGGGCTGC 660  
 QY 661 ACCCTGAATCTTCCCATCAGCCCATCGAGACCGTGGCGCGTGAAGCTGAAGCCCGCATG 720  
 Db 661 ACCCTGAATCTTCCCATCAGCCCATCGAGACCGTGGCGCGTGAAGCTGAAGCCCGCATG 720  
 QY 721 GACGCGCCAGGTGAAGACTGGCCCTGACCGAGAGAGATCAAGGCGCTGACCGCC 780  
 Db 721 GACGCGCCAGGTGAAGACTGGCCCTGACCGAGAGAGATCAAGGCGCTGACCGCC 780  
 QY 781 ATCTCGAGGAGATGGAAGAGGCGCAAGATCACCAAGATCGGCGCGAGAACCCCTAC 840  
 Db 781 ATCTCGAGGAGATGGAAGAGGCGCAAGATCACCAAGATCGGCGCGAGAACCCCTAC 840  
 QY 841 AACGCGCGCTTTCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
 Db 841 AACGCGCGCTTTCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
 QY 901 TTCGCGAGCTGAACAAAGCGACCCAGGACTTCTGGAGGTGACGCTGGGCGATCCCGCCAC 960  
 Db 901 TTCGCGAGCTGAACAAAGCGACCCAGGACTTCTGGAGGTGACGCTGGGCGATCCCGCCAC 960  
 QY 961 CCGCGCGCTGAAG 1020  
 Db 961 CCGCGCGCTGAAG 1020  
 QY 1021 AGCGTGGCGCTGAG 1080  
 Db 1021 AGCGTGGCGCTGAG 1080  
 QY 1081 AACGAG 1140  
 Db 1081 AACGAG 1140  
 QY 1141 CCGAGATCTTCCAG 1200  
 Db 1141 CCGAGATCTTCCAG 1200  
 QY 1201 GAGATCGTATCTACAGGCGCCCTGTGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260

Db 1201 GAGATCGTATCTACAGGCGCCCTGTGATGTTGGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
 QY 1261 CCGCCCAAGATCGAGAGAGTGGCGCAAGCACTGCTGCGCTGGGGCTTCAACACCCCGGAC 1320  
 Db 1261 CCGCCCAAGATCGAGAGAGTGGCGCAAGCACTGCTGCGCTGGGGCTTCAACACCCCGGAC 1320  
 QY 1321 AAGAGACACAG 1380  
 Db 1321 AAGAGACACAG 1380  
 QY 1381 TGGACCTGTGAG 1440  
 Db 1381 TGGACCTGTGAG 1440  
 QY 1441 AAGCTGTGGCGAG 1500  
 Db 1441 AAGCTGTGGCGAG 1500  
 QY 1501 CTGTGAG 1560  
 Db 1501 CTGTGAG 1560  
 QY 1561 GCGAGCTGAG 1620  
 Db 1561 GCGAGCTGAG 1620  
 QY 1621 TACGACCCAG 1680  
 Db 1621 TACGACCCAG 1680  
 QY 1681 TACGAGATCTACAG 1740  
 Db 1681 TACGAGATCTACAG 1740  
 QY 1741 ACCGCGCGAG 1800  
 Db 1741 ACCGCGCGAG 1800  
 QY 1801 AGCATCTGATCTGGGCGAG 1860  
 Db 1801 AGCATCTGATCTGGGCGAG 1860  
 QY 1861 GAGACCTGTGTGAG 1920  
 Db 1861 GAGACCTGTGTGAG 1920  
 QY 1921 ACCGCGCGAG 1980  
 Db 1921 ACCGCGCGAG 1980  
 QY 1981 ACCTTCTACGAG 2040  
 Db 1981 ACCTTCTACGAG 2040  
 QY 2041 ACCGAGCGGCGAG 2100  
 Db 2041 ACCGAGCGGCGAG 2100  
 QY 2101 CTGAG 2160  
 Db 2101 CTGAG 2160  
 QY 2161 AGCAG 2220  
 Db 2161 AGCAG 2220  
 QY 2221 AACAG 2280  
 Db 2221 AACAG 2280  
 QY 2281 CACAG 2340  
 Db 2281 CACAG 2340



```

1621 TACGACCCAGCAAGACCTGTGGCCGAGATCCAGAACGAGGCGACCACTGGACC 1680
1681 TACGAGATCTACGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGC 1740
1681 TACGAGATCTACGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGC 1740
1741 ACGCCGCCACCAACGAGCTGAAGACGCTGACCGAGGCGGCTGCGAAGATCGCCATGGAG 1800
1741 ACGCCGCCACCAACGAGCTGAAGACGCTGACCGAGGCGGCTGCGAAGATCGCCATGGAG 1800
1801 AGCATCGTGTATCTGGGGCAAGACCCCAAGTTCCCGCTGCCCATCCAGAGGAGACCTGG 1860
1801 AGCATCGTGTATCTGGGGCAAGACCCCAAGTTCCCGCTGCCCATCCAGAGGAGACCTGG 1860
1861 GAGACCTGTGTGACCGACTACTTGGGACGACCTGATCCCGGAGTGGGATTCGTGAAC 1920
1861 GAGACCTGTGTGACCGACTACTTGGGAGGCGACCTGATCCCGGAGTGGGATTCGTGAAC 1920
1921 ACCCCGCCCTCGTGTGAAGCTGTGTGACGCTGGAAGAGGAGCCCATCATCGGCGCGGAG 1980
1921 ACCCCGCCCTCGTGTGAAGCTGTGTGACGCTGGAAGAGGAGCCCATCATCGGCGCGGAG 1980
1981 ACCTTCTAGTGTGAGCGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGGCTACGTG 2040
1981 ACCTTCTAGTGTGAGCGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGGCTACGTG 2040
2041 ACCGACCGGGCGCGGACAGATCGTGTGACGCTGAGACGACCAACCAAGAGACCGGAG 2100
2041 ACCGACCGGGCGCGGACAGATCGTGTGACGCTGAGACGACCAACCAAGAGACCGGAG 2100
2101 CTGCGAGCCATCAGCTGCGCTGACGACGAGCGGCGGAGGTTGAACATCGTGACCGAC 2160
2101 CTGCGAGCCATCAGCTGCGCTGACGACGAGCGGCGGAGGTTGAACATCGTGACCGAC 2160
2161 AGCCAGTACGCCCTGCGCATCATCCAGGCCAGCCCGGCAAGAGCGAGCGAGCTGGTG 2220
2161 AGCCAGTACGCCCTGCGCATCATCCAGGCCAGCCCGGCAAGAGCGAGCGAGCTGGTG 2220
2221 AACCAGATCATCGAGCTGATCAAGAGGAGGAGGTTGACCTGAGCTGGTGCGCCGCC 2280
2221 AACCAGATCATCGAGCTGATCAAGAGGAGGAGGTTGACCTGAGCTGGTGCGCCGCC 2280
2281 CACAGGGCATCGGCGGCAACGAGCAGATCGACAGCTGTGTGAGCAAGGGGCATCCGCAAG 2340
2281 CACAGGGCATCGGCGGCAACGAGCAGATCGACAGCTGTGTGAGCAAGGGGCATCCGCAAG 2340
2341 GTGCTGTTCTGAGCGGCTGATCGATGGCGGCTGATCTACGAGTACATGAGGACCTG 2400
2341 GTGCTGTTCTGAGCGGCTGATCGATGGCGGCTGATCTACGAGTACATGAGGACCTG 2400
2401 TACGTGGGAGCGGCGGCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2457
2401 TACGTGGGAGCGGCGGCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2457

```

```

RESULT 3
US-10-190-435-45
; Sequence 45, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Esreilita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: PPI8133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190.435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ ID NO 45
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-435-45

Query Match
Best Local Similarity 98.7%; Score 2430.2; DB 14; Length 2457;
Matches 2448; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY 7 GCCACCATGCGGAGGCGCATGACCGCCAGCGCCAGCATCTCTGATGAGGCGCAGC 66
DB 1 GCCACCATGCGGAGGCGCATGACCGCCAGCGCCAGCATCTCTGATGAGGCGCAGC 60
QY 67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGCCACATC 126
DB 61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGCCACATC 120
QY 127 GCCCGCAACTGCGGCGGCCCCCGCAGAGAGGGCTGTGGAAGTGCGGCAAGGAGGCCAC 186
DB 131 GCCCGCAACTGCGGCGGCCCCCGCAGAGAGGGCTGTGGAAGTGCGGCAAGGAGGCCAC 180
QY 187 CAGATGAAGACTGCAACCGAGCGCCAGGCCAACTTTTCGCGAGGACTGTGGCTTCCCG 246
DB 181 CAGATGAAGACTGCAACCGAGCGCCAGGCCAACTTTTCGCGAGGACTGTGGCTTCCCG 240
QY 247 CAGGCGAAGGCGCGGAGTTTCCCGCAGCGAGCAAGAACCGCGCCAGCCCGCAGCGCGC 306
DB 241 CAGGCGAAGGCGCGGAGTTTCCCGCAGCGAGCAAGAACCGCGCCAGCCCGCAGCGCGC 300
QY 307 GAGCTGCAAGTGTGCGGCGCAAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGC 366
DB 301 GAGCTGCAAGTGTGCGGCGCAAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGC 360
QY 367 CTGAACCTTCCCGCAGATACCCCTGTGCGAGCGCCCTCTGTGAGCATCAAGGTGGCGCGC 426
DB 361 CTGAACCTTCCCGCAGATACCCCTGTGCGAGCGCCCTCTGTGAGCATCAAGGTGGCGCGC 420
QY 427 CAGATCAAGAGAGGCGCTGTGCAACCGGCGCGAGCAACCGTGTGTGAGGAGATGAGC 486
DB 421 CAGATCAAGAGAGGCGCTGTGCAACCGGCGCGAGCAACCGTGTGTGAGGAGATGAGC 480
QY 487 CTGCGCGCAAGTGAAGCCCAAGATGATCGGCGGATCGGGGGCTTCAATCAAGGTGGCGC 546
DB 481 CTGCGCGCAAGTGAAGCCCAAGATGATCGGCGGATCGGGGGCTTCAATCAAGGTGGCGC 540
QY 547 CAGTACGACGAGATCTGTGATCGAGATCTGCGGCAAGAGGCCATCGGCAACCGTGTGTATC 606
DB 541 CAGTACGACGAGATCTGTGATCGAGATCTGCGGCAAGAGGCCATCGGCAACCGTGTGTATC 600
QY 607 GCGCCCAACCGCGTGAACATCATCGGCGCGCAACATCTGACCCAGCTGGGCTGCAACCTG 666
DB 601 GCGCCCAACCGCGTGAACATCATCGGCGCGCAACATCTGACCCAGCTGGGCTGCAACCTG 660
QY 667 AACTTCCCGCATCAGCCCGCATCGAGACCGTGTGCGTGAAGCTGAAAGCCCGGATGAGCGCG 726
DB 661 AACTTCCCGCATCAGCCCGCATCGAGACCGTGTGCGTGAAGCTGAAAGCCCGGATGAGCGCG 720
QY 727 CCCAAGGTGAAGCAGTGGCGCCCTTGAAGCGAGGAGGATCAAGCGCCCTGACCGCCATCTGC 786
DB 721 CCCAAGGTGAAGCAGTGGCGCCCTTGAAGCGAGGAGGATCAAGCGCCCTGACCGCCATCTGC 780
QY 787 GAGGAGATGAGAGAGGAGGCGCAAGATCAACAGATCGGCGCGCGAGAACCCCTTACAAACCC 846
DB 781 GAGGAGATGAGAGAGGAGGCGCAAGATCAACAGATCGGCGCGCGAGAACCCCTTACAAACCC 840
QY 847 CCCGTGTTCCGCTCAAGAGAGAGGAGCAGCACCAAGTGGCGCAAGGTGTGAGTCTCCGC 906
DB 841 CCCGTGTTCCGCTCAAGAGAGAGGAGCAGCACCAAGTGGCGCAAGGTGTGAGTCTCCGC 900
QY 907 GAGCTGAAGAGCGCACCCGAGGACTTCTGTGGAGGCTGACAGCTGGGGCATTCGCCCAACCCGCC 966

```



247 QY CAGGCAAGCGCGAGTTCCTCCAGCAGAGAAACCGCGCCAAACAGCCCAACCAAGCCGC 306  
 241 Db CAGGCAAGCGCGAGTTCCTCCAGCAGAGAAACCGCGCCAAACAGCCCAACCAAGCCGC 300  
 307 QY GAGCTGAGGTGCGCGGAGCAAAACCCCGAGAGAGCGCGCCCGAGCGCCAGAGGAC 366  
 301 Db GAGCTGAGGTGCGCGGAGCAAAACCCCGAGAGAGCGCGCCCGAGCGCCAGAGGAC 360  
 367 QY CTGAATCTTCCCGCAGATCAACCTGTGGCAGCGCCCTGTGTAGCATCAAGGTGGCGGC 426  
 361 Db CTGAATCTTCCCGCAGATCAACCTGTGGCAGCGCCCTGTGTAGCATCAAGGTGGCGGC 420  
 427 QY CAGATCAAGAGAGCGCTGTGAGCAACCGCGCCGAGCAACACCTGTGTGAGAGATGAC 486  
 421 Db CAGATCAAGAGAGCGCTGTGAGCAACCGCGCCGAGCAACACCTGTGTGAGAGATGAC 480  
 487 QY CTGCGCGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGGCTTCATCAAGGTGCGC 546  
 481 Db CTGCGCGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGGCTTCATCAAGGTGCGC 540  
 547 QY CAGTACGACAGATCTGTGATCGAGATCTGCGCAAGAGGCCATCGGCACCGTGTGATC 606  
 541 Db CAGTACGACAGATCTGTGATCGAGATCTGCGCAAGAGGCCATCGGCACCGTGTGATC 600  
 607 QY GCGCCACCGCGTGAACATCATCGCGCCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 666  
 601 Db GCGCCACCGCGTGAACATCATCGCGCCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 660  
 667 QY AACTTCCCATCAGCCCATCGAGACCGTGGCGGCTGAAGCTGAAGCCCGGCATGAGCGC 726  
 661 Db AACTTCCCATCAGCCCATCGAGACCGTGGCGGCTGAAGCTGAAGCCCGGCATGAGCGC 720  
 727 QY CCGAAGGTGAGCAGTGGCCCTCAGCGAGAGAGATCAAGGCCCTGACCGCATCTGC 786  
 721 Db CCGAAGGTGAGCAGTGGCCCTCAGCGAGAGAGATCAAGGCCCTGACCGCATCTGC 780  
 787 QY GAGGAGTGAAGAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTCAACACC 846  
 781 Db GAGGAGTGAAGAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTCAACACC 840  
 847 QY CCGTGTTCGCGATCAAGAAAGAGAGACCAAGTGGCGCAGCTGGTGGATTCGCG 906  
 841 Db CCGTGTTCGCGATCAAGAAAGAGAGACCAAGTGGCGCAGCTGGTGGATTCGCG 900  
 907 QY GAGCTGAACAAAGCGCACCCAGACTTCTGGAGGTGCACTGGGCTGCGCCATCCCGCCGCC 966  
 901 Db GAGCTGAACAAAGCGCACCCAGACTTCTGGAGGTGCACTGGGCTGCGCCATCCCGCCGCC 960  
 967 QY GGCCTGAAGAGAGAGAGCGTGACCGTGTGAGAGTGGCGAGCGCTACTTCAAGCTG 1026  
 961 Db GGCCTGAAGAGAGAGAGCGTGACCGTGTGAGAGTGGCGAGCGCTACTTCAAGCTG 1020  
 1027 QY CCGCTGGAGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCGAGCATCAACAAAGAG 1086  
 1021 Db CCGCTGGAGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCGAGCATCAACAAAGAG 1080  
 1087 QY ACCCGCGGCATCCGCTACAGTACAAAGTGTGCTGCCCGAGGTGAGAGGCGAGCCCGAC 1146  
 1081 Db ACCCGCGGCATCCGCTACAGTACAAAGTGTGCTGCCCGAGGTGAGAGGCGAGCCCGAC 1140  
 1147 QY ATCTTCCAGAGCAGATGACCAAGATCTGTGAGAGCCCTTCCGCGCCGCAACCCCGAGATC 1206  
 1141 Db ATCTTCCAGAGCAGATGACCAAGATCTGTGAGAGCCCTTCCGCGCCGCAACCCCGAGATC 1200  
 1207 QY GTGATCTACCA-----GCGCCCGCTGTAGTGGGAGGCACTTGGAGATGGCCAGCAC 1260  
 1201 Db GTGATCTACCA-----GCGCCCGCTGTAGTGGGAGGCACTTGGAGATGGCCAGCAC 1260  
 1261 QY CCGCGCAAGATGAGAGCTGGCAGCACCTGTGCGTGGGCTTCAACACCCCGCAC 1320  
 1261 Db CCGCGCAAGATGAGAGCTGGCAGCACCTGTGCGTGGGCTTCAACACCCCGCAC 1320

1321 QY AAGAGCAACAG 1380  
 1321 Db AAGAGCAACAG 1380  
 1381 QY TGGACCGCTGAG 1440  
 1381 Db TGGACCGCTGAG 1440  
 1441 QY AAGTGTGGGCAAGTGAAGTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 1441 Db AAGTGTGGGCAAGTGAAGTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 1501 QY CTGTGCAAGTGTGTGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
 1501 Db CTGTGCAAGTGTGTGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
 1561 QY GCGAGCTGAG 1620  
 1561 Db GCGAGCTGAG 1620  
 1621 QY TACGACCCAGCAAG 1680  
 1621 Db TACGACCCAGCAAG 1680  
 1681 QY TACAGATCTACAG 1740  
 1681 Db TACAGATCTACAG 1740  
 1741 QY ACCGCGCCACCAAG 1800  
 1741 Db ACCGCGCCACCAAG 1800  
 1801 QY AGCATCTGTGATCTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860  
 1801 Db AGCATCTGTGATCTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860  
 1861 QY GAGACTGTGAG 1920  
 1861 Db GAGACTGTGAG 1920  
 1921 QY ACCCGCGCGCTGTGATCTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
 1921 Db ACCCGCGCGCTGTGATCTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
 1981 QY ACCTTCTAGTGGAG 2040  
 1981 Db ACCTTCTAGTGGAG 2040  
 2041 QY ACCGACCGCGCGCGAG 2100  
 2041 Db ACCGACCGCGCGCGAG 2100  
 2101 QY CTGACGCGCATCCAGTGGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160  
 2101 Db CTGACGCGCATCCAGTGGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160  
 2161 QY AGCGAGTACGCGCTGTGAG 2220  
 2161 Db AGCGAGTACGCGCTGTGAG 2220  
 2221 QY AACGAGATCTAG 2280  
 2221 Db AACGAGATCTAG 2280  
 2281 QY CACAGGGGATCGCGCGCAAG 2340  
 2281 Db CACAGGGGATCGCGCGCAAG 2340  
 2341 QY GTGCTGTTCAG 2400  
 2341 Db GTGCTGTTCAG 2400  
 2401 QY TACGTGGCGAGCGCGCGCTTAGGATCGATTAAAGGCTTCCCGGGGCTTAGCACCGGT 2457

Db 2401 TACGTGGCAGCGCGCCCTAGGATCGATTAAAGCTTTCCCGGGCTAGCACCGGT 2457

RESULT 5  
US-10-190-435-43  
; Sequence 43, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR NEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: P18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 43  
; LENGTH: 2445  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: p2Pol.opt.YMMW\_c  
US-10-190-435-43

Query Match 98.4%; Score 2422.6; DB 14; Length 2445;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 2441; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

QY 7 GCACCATGCGCGAGGCGCATGAGCCAGGCGCACACGAGCGGCAACATCTGATCGACGCGAGC 66  
Db 1 GCACCATGCGCGAGGCGCATGAGCGGCGCACACGAGCGGCGCACATCTGATCGACGCGAGC 60

QY 67 AACTTCAAGGGCCCCAAGCGCATCAAGTCTTCAACTGCGGCAAGGCGGCACATC 126  
Db 61 AACTTCAAGGGCCCCAAGCGCATCAAGTCTTCAACTGCGGCAAGGCGGCACATC 120

QY 127 GCGCGCAACTGCGCGCGCCCCCGCGAGAGAGGCTGCTGGAAGTGCGGCAAGAGGGCCAC 186  
Db 121 GCGCGCAACTGCGCGCGCCCCCGCGAGAGAGGCTGCTGGAAGTGCGGCAAGAGGGCCAC 180

QY 187 CAGATGAAGGACTGCAACGAGCGCGCAGGCGCAACTTCTTCCGCGAGGACCTGGCCCTTCCC 246  
Db 181 CAGATGAAGGACTGCAACGAGCGCGCAGGCGCAACTTCTTCCGCGAGGACCTGGCCCTTCCC 240

QY 247 CAGGCGAAGCGCGCGAGTTCCCGAGGAGAGAACCGCGCCAAACAGCCCCCAGCGCGC 306  
Db 241 CAGGCGAAGCGCGCGAGTTCCCGAGGAGAGAACCGCGCCAAACAGCCCCCAGCGCGC 300

QY 307 GAGCTGAGGTGCGGGGAGCAACCCCGCGAGGCGGCGCGCGCGCGCGCGCGCGCGCGC 366  
Db 301 GAGCTGAGGTGCGGGGAGCAACCCCGCGAGGCGGCGCGCGCGCGCGCGCGCGCGCGC 360

QY 367 CTGAATCTCCCGAGATCACTCTGTGCGAGCGCGCCCCCTGTGAGCATCAAGGTGGCGCGC 426  
Db 361 CTGAATCTCCCGAGATCACTCTGTGCGAGCGCGCCCCCTGTGAGCATCAAGGTGGCGCGC 420

QY 427 CAGATCAAGGAGCGCGCTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 486  
Db 421 CAGATCAAGGAGCGCGCTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 480

QY 487 CTGCGCGGCAAGTGAAGCGCGCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGC 546  
Db 481 CTGCGCGGCAAGTGAAGCGCGCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGC 540

QY 547 CAGTACGACAGATCTCTGATCGAGATCTGCGCGCAAGAGGCGCATCGCGCGCTGATC 606  
Db 541 CAGTACGACAGATCTCTGATCGAGATCTGCGCGCAAGAGGCGCATCGCGCGCTGATC 600

QY 607 GSCCCACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 666  
Db 601 GSCCCACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 660

QY 667 AACTTCCCATCAGCGCGCATCGAGACCGTGCCTGAGCTGAGCTGAGCGCGCATGAGCGC 726  
Db 661 AACTTCCCATCAGCGCGCATCGAGACCGTGCCTGAGCTGAGCTGAGCGCGCATGAGCGC 720

QY 727 CCAAGGTGAAGCAGTGGCCCTTCAACGAGGAGAGATCAAGGCGCTTGAACCGCATCTGC 786  
Db 721 CCAAGGTGAAGCAGTGGCCCTTCAACGAGGAGAGATCAAGGCGCTTGAACCGCATCTGC 780

QY 787 GAGGAGATGAGAGAGGAGGCAAGATCAACCAAGATCGGCGCGAGACCCCTTCAACAC 846  
Db 781 GAGGAGATGAGAGAGGAGGCAAGATCAACCAAGATCGGCGCGAGACCCCTTCAACAC 840

QY 847 CCGGTGTTGGCCATCAAG 906  
Db 841 CCGGTGTTGGCCATCAAG 900

QY 907 GAGCTGAACAGCGCGCACCGAGACTTCTGGAGGTGAGCTGGGCAATCCCGCACCGCGC 966  
Db 901 GAGCTGAACAGCGCGCACCGAGACTTCTGGAGGTGAGCTGGGCAATCCCGCACCGCGC 960

QY 967 GGCCTGAAG 1026  
Db 961 GGCCTGAAG 1020

QY 1027 CCGCTGAG 1086  
Db 1021 CCGCTGAG 1080

QY 1087 ACCCGCGCATCCGCTACAGTCAACAGTGTGCTGCGCGAGGCTGGAAGGAGAGAGAG 1146  
Db 1081 ACCCGCGCATCCGCTACAGTCAACAGTGTGCTGCGCGAGGCTGGAAGGAGAGAGAG 1140

QY 1147 ATCTTCCAG 1206  
Db 1141 ATCTTCCAG 1200

QY 1207 GTGATCTACAGAGCGCGCTTCTGATGAGTGGGCTACGAGTGGGAGAGAGAGAGAGAG 1266  
Db 1201 GTGATCTACAGAGCGCGCTTCTGATGAGTGGGAGAGAGAGAGAGAGAGAGAGAGAG 1260

QY 1267 AAGATCGAGGAG 1326  
Db 1261 AAGATCGAGGAG 1320

QY 1327 CACCAAG 1386  
Db 1321 CACCAAG 1374

QY 1387 GTGAG 1446  
Db 1375 GTGAG 1434

QY 1447 GTGGCAAG 1506  
Db 1435 GTGGCAAG 1494

QY 1507 AAGCTGTGCGCGCGCGCAAGGCGCTTCAACGAGATCGTGGCGCTTGAACCGAGAGAG 1566  
Db 1495 AAGCTGTGCGCGCGCGCAAGGCGCTTCAACGAGATCGTGGCGCTTGAACCGAGAGAG 1554

QY 1567 CTGGAG 1626  
Db 1555 CTGGAG 1614

QY 1627 CCGAG 1686  
Db 1615 CCGAG 1674

QY 1687 ATCTACAGGAG 1746











RESULT 8  
 US-10-190-435-10  
 : Sequence 10, Application US/10190435  
 : Publication No. US20030143248A1  
 : GENERAL INFORMATION:  
 : APPLICANT: ZUR MESEDE, Jan  
 : APPLICANT: BARNETT, Susan W.  
 : APPLICANT: LIAN, Ying  
 : APPLICANT: ENGELBRECHT, Susan  
 : APPLICANT: VAN RENSBERG, Batrelita J.  
 : TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
 : TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
 : FILE REFERENCE: P218133.003 / 2302-18133  
 : CURRENT APPLICATION NUMBER: US/10/190,435  
 : NUMBER OF SEQ ID NOS: 319  
 : SOFTWARE: PatentIn Ver. 2.0  
 : SEQ ID NO 10  
 : LENGTH: 3930  
 : TYPE: DNA  
 : ORGANISM: Artificial Sequence  
 : FEATURE:  
 : OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutAtt\_C  
 : US-10-190-435-10

Query Match 98.0%; Score 2414; DB 14; Length 3930;  
 Best Local Similarity 99.5%; Pred. No. 0;  
 Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;

Qy	14	TGGCGAGGCGCATGAGCCAGGCGCCACGAGCGCCAACTCCCTGATGCGAGCGCAACTTCA	73
Db	1487	TGCGCGAGGCGCATGAGCCAGGCGCCACGAGCGCCAACTCCCTGATGCGAGCGCAACTTCA	1546
Qy	74	AGGCGCCCAAGCGCATCATCAAGTGTCTCACTCGGCGAGGAGGCGCCACATCGCCGCA	133
Db	1547	AGGCGCCCAAGCGCATCATCAAGTGTCTCACTCGGCGAGGAGGCGCCACATCGCCGCA	1606
Qy	134	ACTGCGCGCGCCCGCGAAGAGGCGCTGCTGGAAGTGGCGAAGAGGCGCCACAGATGA	193
Db	1607	ACTGCGCGCGCCCGCGAAGAGGCGCTGCTGGAAGTGGCGAAGAGGCGCCACAGATGA	1666
Qy	194	AGGACTGACCGAGGCGCGAGCGCAACTTCTTCGCGAGGAGCTGGCTTCCCCCAGGGCA	253
Db	1667	AGGACTGACCGAGGCGCGAGCGCAACTTCTTCGCGAGGAGCTGGCTTCCCCCAGGGCA	1726
Qy	254	AGGCGCGGAGTTCGCCAGCGAGCAGAACCCGCGCCAAAGCCGCGCCAGCGCGAGTGC	313
Db	1727	AGGCGCGGAGTTCGCCAGCGAGCAGAACCCGCGCCAAAGCCGCGCCAGCGCGAGTGC	1786
Qy	314	AGGTGCGCGGACAAACCCCGCGAGCGAGGCGCGCGCGCGAGCGCGCGAGTGC	373
Db	1787	AGGTGCGCGGACAAACCCCGCGAGCGAGGCGCGCGCGCGAGCGCGCGAGTGC	1846
Qy	374	TCCCCCAGATCACCTGTGGAGCGCCCTTGTGAGCATCAAGTGGCGGCGCGAGATCA	433
Db	1847	TCCCCCAGATCACCTGTGGAGCGCCCTTGTGAGCATCAAGTGGCGGCGCGAGATCA	1906
Qy	434	AGGAGGCGCTGTGACACCGCGCGCGAGCAGAACCGTGTGAGAGATGAGCTGCGCG	493
Db	1907	AGGAGGCGCTGTGACACCGCGCGCGAGCAGAACCGTGTGAGAGATGAGCTGCGCG	1966
Qy	494	GCAAGTGGAGCCCAAGATGATCGGCGGATCGGCGGCTTCAAGTGGCGCGAGTACG	553
Db	1967	GCAAGTGGAGCCCAAGATGATCGGCGGATCGGCGGCTTCAAGTGGCGCGAGTACG	2026
Qy	554	ACCAGATCTGTGATGAGATCTCGGCGAAGAGGCGCATCGGACCGTGTGATCGGCGCA	613
Db	2027	ACCAGATCTGTGATGAGATCTCGGCGAAGAGGCGCATCGGACCGTGTGATCGGCGCA	2086
Qy	614	CCCCCGTGAACATCATCGGCGCGAGATGCTGACCCAGCTGGCTGACCCGACTTCC	673
Db	2087	CCCCCGTGAACATCATCGGCGCGAGATGCTGACCCAGCTGGCTGACCCGACTTCC	2146

Qy	674	CCATCAGCCCATGAGACCGTGCCTGAAGCTGAAGCCCGGCGCATGAGCGGCCCAAGG	733
Db	2147	CCATCAGCCCATGAGACCGTGCCTGAAGCTGAAGCCCGGCGCATGAGCGGCCCAAGG	2206
Qy	734	TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTGCGAGAGA	793
Db	2207	TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTGCGAGAGA	2266
Qy	794	TGGAGAGGAGGCGAGATCAACCAAGATCGGCCCGGAGAACCCCTACAAACCCCGCTGT	853
Db	2267	TGGAGAGGAGGCGAGATCAACCAAGATCGGCCCGGAGAACCCCTACAAACCCCGCTGT	2326
Qy	854	TGCGCATCAAGAAGAGGAGCAGCACCAGTGGCGCAAGTGGTGGAGCTTCCGCGAGCTGA	913
Db	2327	TGCGCATCAAGAAGAGGAGCAGCACCAGTGGCGCAAGTGGTGGAGCTTCCGCGAGCTGA	2386
Qy	914	ACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGAGTGGGCGATCCCCCAGCGGCTGA	973
Db	2387	ACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGAGTGGGCGATCCCCCAGCGGCTGA	2446
Qy	974	AGAAGAAGAGGCGTGAACCGTGTGAGCTGGGCGAGCGCTTCTCAGCGTGGCGCTTCG	1033
Db	2447	AGAAGAAGAGGCGTGAACCGTGTGAGCTGGGCGAGCGCTTCTCAGCGTGGCGCTTCG	2506
Qy	1034	ACGAGGACTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACAGAGAGCCCGG	1093
Db	2507	ACGAGGACTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACAGAGAGCCCGG	2566
Qy	1094	GCATCGCTTACCAAGTACAAAGTGTGCGCCAGAGGCTGGAAGGCGAGCCCGAGCATCTTCC	1153
Db	2567	GCATCGCTTACCAAGTACAAAGTGTGCGCCAGAGGCTGGAAGGCGAGCCCGAGCATCTTCC	2626
Qy	1154	AGAGCAGCATGACCAAGTACCTGGAGCGCTTCCGCGCGCGCAACCCCGAGATCTGATCT	1213
Db	2627	AGAGCAGCATGACCAAGTACCTGGAGCGCTTCCGCGCGCGCAACCCCGAGATCTGATCT	2686
Qy	1214	ACAGGCGCGCGCTGACCTGGAGCGAGCTGGAGTGGCGAGCAGCAGCGCGCGCAAGATCG	1273
Db	2687	ACAGGCGCGCGCTGACCTGGAGCGAGCTGGAGTGGCGAGCAGCAGCGCGCGCAAGATCG	2746
Qy	1274	AGGAGCTGCGCAAGCAGCTGCTGGCGGCTTCAACCCCGCGCAAGAGCAGCAGCA	1333
Db	2747	AGGAGCTGCGCAAGCAGCTGCTGGCGGCTTCAACCCCGCGCAAGAGCAGCAGCA	2806
Qy	1334	AGAGCGCGCGCTTCTGTGGATGGCTAGAGCTGCACCCCGAGAGTGGACCGTGCAGC	1393
Db	2807	AGAGCGCGCGCTTCTGTGGATGGCTAGAGCTGCACCCCGAGAGTGGACCGTGCAGC	2860
Qy	1394	CCATCAGCTGCGCGAGAGAGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGGCA	1453
Db	2861	CCATCAGCTGCGCGAGAGAGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGGCA	2920
Qy	1454	AGTGAATGGGCGAGCAGATCTACCCCGGCGATCAAGTGGCGCGAGCTGTGCGAGCTGC	1513
Db	2921	AGTGAATGGGCGAGCAGATCTACCCCGGCGATCAAGTGGCGCGAGCTGTGCGAGCTGC	2980
Qy	1514	TGCGCGCGCGCAAGGCGCTGACCGCATCTGCGCGCTTGAACCGAGGCGCGAGCTGGAGC	1573
Db	2981	TGCGCGCGCGCAAGGCGCTGACCGCATCTGCGCGCTTGAACCGAGGCGCGAGCTGGAGC	3040
Qy	1574	TGCGCGAGAACCGCGAGATCTTGGCGAGCGCGCTGCAACCGAGGCGCGAGCTGGAGC	1633
Db	3041	TGCGCGAGAACCGCGAGATCTTGGCGAGCGCGCTGCAACCGAGGCGCGAGCTGGAGC	3100
Qy	1634	AGGACTGGTGGCGGAGATCCAGAGCAGGCGCACGACAGTGGAGCTTACACAGATCTACC	1693
Db	3101	AGGACTGGTGGCGGAGATCCAGAGCAGGCGCACGACAGTGGAGCTTACACAGATCTACC	3160
Qy	1694	AGGAGCGCTTCAAGAACCTGAAGACCGGCAAGTGAACCAAGATGGCGCACCGGCCACACCA	1753
Db	3161	AGGAGCGCTTCAAGAACCTGAAGACCGGCAAGTGAACCAAGATGGCGCACCGGCCACACCA	3220

1754 ACACCTGAGCAGCTGACCGAGCGCTGCAGAGATCCCATGAGAGCATCGTGATCT 1813  
3221 ACACCTGAGCAGCTGACCGAGCGCTGCAGAGATCCCATGAGAGCATCGTGATCT 3280  
1814 GGGGCAAGACCCCAAGTTCGCTGCCATCCAGAGGAGACCTGGGAGACTTGTGA 1873  
3281 GGGGCAAGACCCCAAGTTCGCTGCCATCCAGAGGAGACCTGGGAGACTTGTGA 3340  
1874 CCAGTACTGGCAGGCGACCTGGATCCCGAGTGGAGTTCGTGAACACCCCCCGCTGG 1933  
3341 CCAGTACTGGCAGGCGACCTGGATCCCGAGTGGAGTTCGTGAACACCCCCCGCTGG 3400  
1934 TGAAGTGTGTACAGCTGGAAGAGGAGCCATCATCGCGCGGAGAGACTTCTAGTGG 1993  
3401 TGAAGTGTGTACAGCTGGAAGAGGAGCCATCATCGCGCGGAGAGACTTCTAGTGG 3460  
1994 ACAGCGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAGCGCGGCC 2053  
3461 ACAGCGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAGCGCGGCC 3520  
2054 GGCAAGAGATCGTGAAGCTGACCGAGACCAACAGAGAGAGCGAGCTGCGAGCGCATCC 2113  
3521 GGCAAGAGATCGTGAAGCTGACCGAGACCAACAGAGAGAGCGAGCTGCGAGCGCATCC 3580  
2114 AGCTGGCGCTGCAGGAGCAGCGGAGAGGTGAACATCGTGACCGAGCCAGTACGCC 2173  
3581 AGCTGGCGCTGCAGGAGCAGCGGAGAGGTGAACATCGTGACCGAGCCAGTACGCC 3640  
2174 TGGGCAATCATCGAGCCCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2233  
3641 TGGGCAATCATCGAGCCCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3700  
2234 AGCAGCTGATCAAG 2293  
3701 AGCAGCTGATCAAG 3760  
2294 GCGGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGAGAGAGAGAGAGAGAGAGAGAG 2353  
3761 GCGGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGAGAGAGAGAGAGAGAGAGAGAG 3820  
2354 ACAGCATGATGGCGGATCGTGATCTACAGTACATGAGAGAGAGAGAGAGAGAGAGAG 2413  
3821 ACAGCATGATGGCGGATCGTGATCTACAGTACATGAGAGAGAGAGAGAGAGAGAGAG 3880  
2414 GCGGCGCTAGGATCGATTAAGCTTCCCGGGGCTAGCACCGGT 2457  
3881 GCGGCGCTAGGATCGATTAAGCTTCCCGGGGCTAGCACCGGT 3924

RESULT 9  
US-10-150-435-11  
; Sequence 11, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: PP18133.003 / 2302-18133  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 3930  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutina\_C  
US-10-150-435-11

Query Match 98.0%; Score 2414; DB 14; Length 3930;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;  
14 TGGCCGAGGCGCATGAGCAGGCCACACAGGCGCCAAACATCTCTGATGACGCGCAACTTCA 73  
Db 1487 TCGCCGAGGCGCATGAGCAGGCCACACAGGCGCCAAACATCTCTGATGACGCGCAACTTCA 1546  
74 AGGCGCCCAAGCGCATCATCAAGTGTCTTAACTTGGCGCAAGGAGGCGCATCGCCCGCA 133  
Db 1547 AGGCGCCCAAGCGCATCATCAAGTGTCTTAACTTGGCGCAAGGAGGCGCATCGCCCGCA 1606  
134 ACTCCGCGCGCCCGCAAGAGGCTGTGGAAGTGCAGCAAGGAGGCGCCACCATGATGA 193  
Db 1607 ACTCCGCGCGCCCGCAAGAGGCTGTGGAAGTGCAGCAAGGAGGCGCCACCATGATGA 1666  
194 AGGACTGCAACGAGCGCGCATCAAGTGTCTTAACTTTCGCGAGGAGACTTGGCTTCCCCAGGGCA 253  
Db 1667 AGGACTGCAACGAGCGCGCATCAAGTGTCTTAACTTTCGCGAGGAGACTTGGCTTCCCCAGGGCA 1726  
254 AGGCGCGCGAGTTCGCCAGCGAGCAGAAACCGCGCAACAGCCCGCCACAGCGCGGAGTGC 313  
Db 1727 AGGCGCGCGAGTTCGCCAGCGAGCAGAAACCGCGCAACAGCCCGCCACAGCGCGGAGTGC 1786  
314 AGGTGCGCGCGCAACACCCCGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
Db 1787 AGGTGCGCGCGCAACACCCCGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846  
374 TCCCGCAGATCACCTGTGCG 433  
Db 1847 TCCCGCAGATCACCTGTGCG 1906  
434 AGGAGCGCGCTGCTGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493  
Db 1907 AGGAGCGCGCTGCTGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1966  
494 GCAAGTGAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 553  
Db 1967 GCAAGTGAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2026  
554 ACCAGATCCTGATCGAGATCTGCGCGCAAGAGGCGCATCGGACCGTGTCTGATCGGCGCGCA 613  
Db 2027 ACCAGATCCTGATCGAGATCTGCGCGCAAGAGGCGCATCGGACCGTGTCTGATCGGCGCGCA 2086  
614 CCCCCGTGAACATCATTCGCGCGCGCAACATCTGAGCCAGCTGGGCTGCACCTGAACTTCC 673  
Db 2087 CCCCCGTGAACATCATTCGCGCGCGCAACATCTGAGCCAGCTGGGCTGCACCTGAACTTCC 2146  
674 CCATCAGCGCCCATCGAGACCGTGCCTGAAAGCTGAAAGCGCGCATGGAAGCGCCCAAGG 733  
Db 2147 CCATCAGCGCCCATCGAGACCGTGCCTGAAAGCTGAAAGCGCGCATGGAAGCGCCCAAGG 2206  
734 TGAAGCAGTGGCGCGCTGACCGAGGAGAGATCAAGGCGCGCGCGCGCGCGCGCGCGCGCG 793  
Db 2207 TGAAGCAGTGGCGCGCGCTGACCGAGGAGAGATCAAGGCGCGCGCGCGCGCGCGCGCGCG 2266  
794 TGGAGAGAGGCGCAAGATCAACAGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 853  
Db 2267 TGGAGAGAGGCGCAAGATCAACAGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2326  
854 TCGCCATCAAG 913  
Db 2327 TCGCCATCAAG 2386  
914 ACAAGCGCACCGAGACTTCTGGGAGGTGAGCTGGGATCCCCACCGCGCGCGCGCGCTGA 973  
Db 2387 ACAAGCGCACCGAGACTTCTGGGAGGTGAGCTGGGATCCCCACCGCGCGCGCGCTGA 2446  
974 AGAAG 1033  
Db 2447 AGAAG 2506

1034 ACAGGACTTCGCGAAGTACACCGCTTCCACCATCCCGAGCATCAACAAAGAGACCCCG 1093  
 Db ACAGGACTTCGCGAAGTACACCGCTTCCACCATCCCGAGCATCAACAAAGAGACCCCG 2566  
 QY GCATCCGCTACAGTACAAAGTCTGCTCCCGAGGGCTGGAAGGCGACCCAGCATCTTCC 1153  
 Db GCATCCGCTACAGTACAAAGTCTGCTCCCGAGGGCTGGAAGGCGACCCAGCATCTTCC 2626  
 QY AGAGCAGCATGACCAAGATCTTGAGAGCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213  
 Db AGAGCAGCATGACCAAGATCTTGAGAGCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686  
 QY ACCAGGCCCGCTTGAGAGCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1273  
 Db ACCAGGCCCGCTTGAGAGCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2746  
 QY AGAGCTGGCGCAAGCAGCTGCTGCTGGGGCTTCCACCCCGCAACAAAGAGCACCAGA 1333  
 Db AGAGCTGGCGCAAGCAGCTGCTGCTGGGGCTTCCACCCCGCAACAAAGAGCACCAGA 2806  
 QY AGAGGCCCGCTTCTGAGAGCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1393  
 Db AGAGGCCCGCTTCTGAGAGCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2860  
 QY CCATCGAGCTGCGCGAAGAGAGCTGGAACCGTGAACGACATCCAGAAAGCTGGTGGCA 1453  
 Db CCATCGAGCTGCGCGAAGAGAGCTGGAACCGTGAACGACATCCAGAAAGCTGGTGGCA 2920  
 QY AGCTGAATCGGGCAGCAGATCTACCCCGGATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513  
 Db AGCTGAATCGGGCAGCAGATCTACCCCGGATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980  
 QY TGGCGCGCGCAAGGCGCTGACGACATCGTGCCTGACCGAGAGGCGGAGCTGGAGC 1573  
 Db TGGCGCGCGCAAGGCGCTGACGACATCGTGCCTGACCGAGAGGCGGAGCTGGAGC 3040  
 QY TGGCGCGCGCAAGGCGCTGACGACATCGTGCCTGACCGAGAGGCGGAGCTGGAGC 1633  
 Db TGGCGCGCGCAAGGCGCTGACGACATCGTGCCTGACCGAGCGTGTACTACGACCCAGCA 3100  
 QY AGGACCTGTTGGCGAGATCCAGAGCAGGCGCCAGCAGTGGAGCTTACCATCTTCC 1693  
 Db AGGACCTGTTGGCGAGATCCAGAGCAGGCGCCAGCAGTGGAGCTTACCATCTTCC 3160  
 QY AGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCA 1753  
 Db AGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCA 3220  
 QY ACAGCTGAAGCAGCTGACCGAGCCCGTGCAGAAAGATCGCCATGGAGAGATCGTGATCT 1813  
 Db ACAGCTGAAGCAGCTGACCGAGCCCGTGCAGAAAGATCGCCATGGAGAGATCGTGATCT 3280  
 QY GGGGCAAGACCCCAAGTTCGCTGCGCCATCCAGAGGAGACCTGGGAGACCTGGTGA 1873  
 Db GGGGCAAGACCCCAAGTTCGCTGCGCCATCCAGAGGAGACCTGGGAGACCTGGTGA 3340  
 QY CCGACTAGTGCAGGCGCACCTGGATCCCGAGTGGAGGTTCTGTAACACCCCGCCCTGG 1933  
 Db CCGACTAGTGCAGGCGCACCTGGATCCCGAGTGGAGGTTCTGTAACACCCCGCCCTGG 3400  
 QY TGAAGCTGTGTGTAACAGTGAAGAGGAGCCATCATCGCGCGGAGACCTTCTTACGTGG 1993  
 Db TGAAGCTGTGTGTAACAGTGAAGAGGAGCCATCATCGCGCGGAGACCTTCTTACGTGG 3460  
 QY ACAGGCGCGCAACCGCGAGACCAAGATCGGAGGCGGCTAGTCCAGCGCGGCGCC 2053  
 Db ACAGGCGCGCAACCGCGAGACCAAGATCGGAGGCGGCTAGTCCAGCGCGGCGCC 3520  
 QY GGCAGAGATCGTGAGGCTGACCGAGAGCACCACCAAGAGACCGAGCTGAGGCCATCC 2113  
 Db GGCAGAGATCGTGAGGCTGACCGAGAGCACCACCAAGAGACCGAGCTGAGGCCATCC 3580  
 QY AGCTGGCCCTGCGAGGAGCGGAGGAGTGAACATCGTGACCGACGAGCCAGTACGCC 2173

Db 3581 AGCTGGCCCTGCGAGGACGCGGAGCGAGGTGAACATCGTACCGGACGCGAGTACGCC 3640  
 QY 2174 TGGGCATCATCCAGGCGGAGCCCGCAAGAGCGAGGAGCTGGTGAACGAGATCATCG 2233  
 Db 3641 TGGGCATCATCCAGGCGGAGCCCGCAAGAGCGAGGAGCTGGTGAACGAGATCATCG 3700  
 QY 2234 AGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCGCCCGCAACAGGGCATCG 2293  
 Db 3701 AGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCGCCCGCAACAGGGCATCG 3760  
 QY 2294 GCGGCAACGAGCAGATCGAACAAGTGTGAGCAAGGGCATCCGCAAGTGTCTTCTGTG 2353  
 Db 3761 GCGGCAACGAGCAGATCGAACAAGTGTGAGCAAGGGCATCCGCAAGTGTCTTCTGTG 3820  
 QY 2354 ACAGCATGATGCGGCGATCGTATCTACAGTACATGAGCAGCTGTAGTGTGGGAGCG 2413  
 Db 3821 ACAGCATGATGCGGCGATCGTATCTACAGTACATGAGCAGCTGTAGTGTGGGAGCG 3880  
 QY 2414 GCGGCCCTAGGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 2457  
 Db 3881 GCGGCCCTAGGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 3924

RESULT 10  
 US-10-190-435-58  
 ; Sequence 58, Application US/10190435  
 ; Publication No. US20030143248A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ZUR MEGEDE, Jan  
 ; APPLICANT: BARNETT, Susan W.  
 ; APPLICANT: LIAN, Ying  
 ; APPLICANT: ENGELBRECHT, Susan  
 ; APPLICANT: VAN RENSBURG, Estrelita J.  
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
 ; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
 ; FILE REFERENCE: P18133.003 / 2302-18133  
 ; CURRENT APPLICATION NUMBER: US/10/190,435  
 ; NUMBER OF SEQ ID NOS: 319  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 58  
 ; LENGTH: 5184  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: TatRevNefgagCpolIna C  
 US-10-190-435-58

Query Match 98.0%; Score 2414; DB 14; Length 5184;  
 Best Local Similarity 99.5%; Pred. No. 0;  
 Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;

QY 14 TGGCCGAGCCATGAGCGGATCATCAAGTGTCTTCAACTGCGCAAGGAGGCGCATATCGCCCGCA 73  
 Db 2741 TCGCCGAGCCATGAGCGGATCATCAAGTGTCTTCAACTGCGCAAGGAGGCGCATATCGCCCGCA 2800  
 QY 74 AGGCGCCCAAGCGGATCATCAAGTGTCTTCAACTGCGCAAGGAGGCGCATATCGCCCGCA 133  
 Db 2801 AGGCGCCCAAGCGGATCATCAAGTGTCTTCAACTGCGCAAGGAGGCGCATATCGCCCGCA 2860  
 QY 134 ACTGCGCGCCCGCCCGCAAGAGGCTCTGGAAGTGGCGCAAGGAGGCGCCACGAGATGA 193  
 Db 2861 ACTGCGCGCCCGCCCGCAAGAGGCTCTGGAAGTGGCGCAAGGAGGCGCCACGAGATGA 2920  
 QY 194 AGGACTGCAACGAGCGCGCAGGCGCAAGTCTTCCCGGAGGAGCTTCCCGCCAGGCGCA 253  
 Db 2921 AGGACTGCAACGAGCGCGCAGGCGCAAGTCTTCCCGGAGGAGCTTCCCGCCAGGCGCA 2980  
 QY 254 AGGCGCGGAGTTCCTCCAGCGGAGAGAACCGCGCCCAAGAGCCCGCCAGCGCGAGTGC 313  
 Db 2981 AGGCGCGGAGTTCCTCCAGCGGAGAGAACCGCGCCCAAGAGCCCGCCAGCGCGAGTGC 3040

QY 314 AGGTGCGGCGAACAACCCCGCAGCGAGCGCCGCGCCGAGCGCCAGGCGCAACCTGAACT 373  
 Db 3041 AGGTGCGGCGGACAAACCCCGCAGCGAGCGCCGCGCCGAGCGCCAGGCGCAACCTGAACT 3100  
 QY 374 TCCCCAGATACCTGTGGAGCGCCCTCGTGGAGCATCAAGTGGGCGGCCAGATCA 433  
 Db 3101 TCCCCAGATACCTGTGGAGCGCCCTCGTGGAGCATCAAGTGGGCGGCCAGATCA 3160  
 QY 434 AGAGGCCCTGTGGACACCGCGCGCGACGACACCGTGTGGAGAGATGAGCTGCCG 493  
 Db 3161 AGAGGCCCTGTGGACACCGCGCGCGACGACACCGTGTGGAGAGATGAGCTGCCG 3220  
 QY 494 GCAAGTGAAGCCCAAGATGATCGGCGCATCGGCGCTTCATCAAGTGGGCCAGTACG 553  
 Db 3221 GCAAGTGAAGCCCAAGATGATCGGCGCATCGGCGCTTCATCAAGTGGGCCAGTACG 3280  
 QY 554 ACCAGATCCTGTGAGATCTGCGGCAAGAGGCCATCGGCAACCGTGTGATCGGCCCA 613  
 Db 3281 ACCAGATCCTGTGAGATCTGCGGCAAGAGGCCATCGGCAACCGTGTGATCGGCCCA 3340  
 QY 614 CCGCGTGAACATCATCGGCGCAACATGCTGACCGAGCTGGGCTGCACCTGAACTTCC 673  
 Db 3341 CCGCGTGAACATCATCGGCGCAACATGCTGACCGAGCTGGGCTGCACCTGAACTTCC 3400  
 QY 674 CCATGAGCCCATCGAGACCGTGCCTGAAAGCTGAAGCCCGGCATGGACGCCCAAG 733  
 Db 3401 CCATGAGCCCATCGAGACCGTGCCTGAAAGCTGAAGCCCGGCATGGACGCCCAAG 3460  
 QY 734 TGAAGCATGCGCCCTGACCGAGGAGATCAAGGCCCTGACCGCCATCTCGGAGGAGA 793  
 Db 3461 TGAAGCATGCGCCCTGACCGAGGAGATCAAGGCCCTGACCGCCATCTCGGAGGAGA 3520  
 QY 794 TGAAGGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTACACACCCCGGT 853  
 Db 3521 TGAAGGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTACACACCCCGGT 3580  
 QY 854 TCGCCATCAAGAGAGGAGACAGACCAAGTGGCGGAGCTGTGGACTTCGCGAGCTGA 913  
 Db 3581 TCGCCATCAAGAGAGGAGACAGACCAAGTGGCGGAGCTGTGGACTTCGCGAGCTGA 3640  
 QY 914 ACAAGCCACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCGCACCCCGCGGCTGA 973  
 Db 3641 ACAAGCCACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCGCACCCCGCGGCTGA 3700  
 QY 974 AGAAGAGAGAGGCTGACCGTGTGACCGTGTGGAGCGAGCTGAGCTGAGCTGAGCTG 1033  
 Db 3701 AGAAGAGAGAGGCTGACCGTGTGACCGTGTGGAGCGAGCTGAGCTGAGCTGAGCTG 3760  
 QY 1034 ACAGAGACTTCGCAAGTACACCGCTTCACCATCCCGAGCATCAACAGAGACCCCG 1093  
 Db 3761 ACAGAGACTTCGCAAGTACACCGCTTCACCATCCCGAGCATCAACAGAGACCCCG 3820  
 QY 1094 GCATCCGCTACCAAGTACCAAGTGTGCTCCCGAGGCTGGAAGGAGCGCCAGATCTTCC 1153  
 Db 3821 GCATCCGCTACCAAGTACCAAGTGTGCTCCCGAGGCTGGAAGGAGCGCCAGATCTTCC 3880  
 QY 1154 AGAGCAGTACCAAGATCTGAGGCTTCGCGCCCGCAACCCCGAGATCGTATCT 1213  
 Db 3881 AGAGCAGTACCAAGATCTGAGGCTTCGCGCCCGCAACCCCGAGATCGTATCT 3940  
 QY 1214 ACCAGGCCCTGTGAGTGGGAGCGACCTGGAGATCGGCCAGACCCCGCAAGATCG 1273  
 Db 3941 ACCAGGCCCTGTGAGTGGGAGCGACCTGGAGATCGGCCAGACCCCGCAAGATCG 4000  
 QY 1274 AGGAGCTGCGCAAGCATCTGTGGCTGGGCTTCACCATCCCGCAAGAGCACCA 1333  
 Db 4001 AGGAGCTGCGCAAGCATCTGTGGCTGGGCTTCACCATCCCGCAAGAGCACCA 4060  
 QY 1334 AGGAGCCCTCTCTGTGATGGGCTACGAGCTGACCCCGCAAGTGGACCGTGCAGC 1393  
 Db 4061 AGGAGCCCTCTCTGTGATGGGCTACGAGCTGACCCCGCAAGTGGACCGTGCAGC 4114  
 QY 1394 CCATGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGCA 1453

Db 4115 CCATGAGCTGCCGAGAGAGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGCA 4174  
 QY 1454 AGTGAACCTGGGCGACCCAGATCTACCCGCGATCAAGTGGCGCAGCTGTGACAGCTGC 1513  
 Db 4175 AGTGAACCTGGGCGACCCAGATCTACCCGCGATCAAGTGGCGCAGCTGTGACAGCTGC 4234  
 QY 1514 TGGCGGCGCAAGGCCCTGACCGATCGTGCCTGACCCGAGGAGCGAGCTGGAGC 1573  
 Db 4235 TGGCGGCGCAAGGCCCTGACCGATCGTGCCTGACCCGAGGAGCGAGCTGGAGC 4294  
 QY 1574 TGGCGGAGAACCCGAGATCTTGGCGAGCCGCTGCAACCGCTGTACTACGACCCAGCA 1633  
 Db 4295 TGGCGGAGAACCCGAGATCTTGGCGAGCCGCTGCAACCGCTGTACTACGACCCAGCA 4354  
 QY 1634 AGGACCTGGTGGCGGAGATCCAGAGCAGGCCCAACAGTGGGACTTACAGATCTACC 1693  
 Db 4355 AGGACCTGGTGGCGGAGATCCAGAGCAGGCCCAACAGTGGGACTTACAGATCTACC 4414  
 QY 1694 AGGAGCCCTTCAAGAACCTGAGAACCGGCAAGTACCCCAAGATGCGCACCGGCCACCA 1753  
 Db 4415 AGGAGCCCTTCAAGAACCTGAGAACCGGCAAGTACCCCAAGATGCGCACCGGCCACCA 4474  
 QY 1754 ACAGCTGAGAGAGCTGACCGAGCCGCTGCAAGATCGCATGGAGAGCATCGTATCT 1813  
 Db 4475 ACAGCTGAGAGAGCTGACCGAGCCGCTGCAAGATCGCATGGAGAGCATCGTATCT 4534  
 QY 1814 GGGCAAGAACCCCAAGTTCGGCTGCCATCCAGAGGAGACTTGGGAGACTTGGTGA 1873  
 Db 4535 GGGCAAGAACCCCAAGTTCGGCTGCCATCCAGAGGAGACTTGGGAGACTTGGTGA 4594  
 QY 1874 CCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTAACACCCCGCCCTGG 1933  
 Db 4595 CCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTAACACCCCGCCCTGG 4654  
 QY 1934 TGAAGCTGTGTGATCAAGCTGGAGAGGCCATCATCGCGCGCGAGACTTCTAGCTGG 1993  
 Db 4655 TGAAGCTGTGTGATCAAGCTGGAGAGGCCATCATCGCGCGCGAGACTTCTAGCTGG 4714  
 QY 1994 ACGGCGCGCCAAACCGGAGACCCAGATCGGCAAGCGCGCTACGTACCGAGCCGGGCC 2053  
 Db 4715 ACGGCGCGCCAAACCGGAGACCCAGATCGGCAAGCGCGCTACGTACCGAGCCGGGCC 4774  
 QY 2054 GGCAGAGATCGTGAAGCTGACCGAGACCCAAACCAAGAGACCGAGCTGAGGCCATCC 2113  
 Db 4775 GGCAGAGATCGTGAAGCTGACCGAGACCCAAACCAAGAGACCGAGCTGAGGCCATCC 4834  
 QY 2114 AGCTGGCCCTGCAAGACAGCGGAGAGGTGATCTGAGCTGGTGCCTCCAGAGCTACG 2173  
 Db 4835 AGCTGGCCCTGCAAGACAGCGGAGAGGTGATCTGAGCTGGTGCCTCCAGAGCTACG 4894  
 QY 2174 TGGGCATCATCCAGGCCCGCCAGCCGACAGAGAGCGAGCTGGTGAACAGATCATCG 2233  
 Db 4895 TGGGCATCATCCAGGCCCGCCAGCCGACAGAGAGCGAGCTGGTGAACAGATCATCG 4954  
 QY 2234 AGCAGCTGATCAAGAGAGAGAGGTGATCTGAGCTGGTGCCTCCAGAGCTACG 2293  
 Db 4955 AGCAGCTGATCAAGAGAGAGAGGTGATCTGAGCTGGTGCCTCCAGAGCTACG 5014  
 QY 2294 GCGGCAAGAGAGATCGACAGCTGTGAGCAAGGCGATCCGCAAGGTGTCTTCTCTGG 2353  
 Db 5015 GCGGCAAGAGAGATCGACAGCTGTGAGCAAGGCGATCCGCAAGGTGTCTTCTCTGG 5074  
 QY 2354 ACGGCATCGATGGGCGCATCGTATCTACAGATCATGGAGAGCTGTACGTGGGCGAGC 2413  
 Db 5075 ACGGCATCGATGGGCGCATCGTATCTACAGATCATGGAGAGCTGTACGTGGGCGAGC 5134  
 QY 2414 GCGGCCCTAGGATCGATTAAAGCTTCCGGGCTTAGCACCGGT 2457  
 Db 5135 GCGGCCCTAGGATCGATTAAAGCTTCCGGGCTTAGCACCGGT 5178

RESULT 11







```
QY 1814 GGGCAAGACCCCAAGTTCGGCTGCCATCCAGAAAGAGACCTGGGAGACCTGGTGA 1873
Db 4535 GGGCAAGACCCCAAGTTCGGCTGCCATCCAGAAAGAGACCTGGGAGACCTGGTGA 4594
QY 1874 CCGACTACTGGCAGGCGACCTGGATCCCGAGTGGAGTTCTGTGAACACACCCCCCTGG 1933
Db 4595 CCGACTACTGGCAGGCGACCTGGATCCCGAGTGGAGTTCTGTGAACACACCCCCCTGG 4654
QY 1934 TGAAGCTGTGTACCGAGTGGAGAGGAGGCCATCATCTCGCGCGGAGACCTTCTAGTGG 1993
Db 4655 TGAAGCTGTGTACCGAGTGGAGAGGAGGCCATCATCTCGCGCGGAGACCTTCTAGTGG 4714
QY 1994 ACGGCGCGCCCAACCGGAGACCGAGATCGGCAAGGCGGCTAGCTGACCGAGCGGCGCC 2053
Db 4715 ACGGCGCGCCCAACCGGAGACCGAGATCGGCAAGGCGGCTAGCTGACCGAGCGGCGCC 4774
QY 2054 GGCAGAGATCGTGAAGCTGTGACCGAGACCCAGACCCAAACCAAGAGACCGAGCTGAGGCCATCC 2113
Db 4775 GGCAGAGATCGTGAAGCTGTGACCGAGACCCAGACCCAAACCAAGAGACCGAGCTGAGGCCATCC 4834
QY 2114 AGCTGGCCCTGACGACGCGGAGCGGAGGTGACATCGTGACGACGACGACGACGACGACG 2173
Db 4835 AGCTGGCCCTGACGACGCGGAGCGGAGGTGACATCGTGACGACGACGACGACGACGACG 4894
QY 2174 TGGGCATCATCCAGGCCCGAGCCGACCAAGAGCGAGCGAGCTGGTGAACCAAGATCATCG 2233
Db 4895 TGGGCATCATCCAGGCCCGAGCCGACCAAGAGCGAGCGAGCTGGTGAACCAAGATCATCG 4954
QY 2234 AGCAGCTGATCAAGAGAGAGGTGTACCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2293
Db 4955 AGCAGCTGATCAAGAGAGAGGTGTACCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 5014
QY 2294 GGGGCAACGAGCAGATCGACAACTGGTGAAGAGGGGATCCGCAAGGTGCTGTTCCTGG 2353
Db 5015 GGGGCAACGAGCAGATCGACAACTGGTGAAGAGGGGATCCGCAAGGTGCTGTTCCTGG 5074
QY 2354 ACGGCATCGATGGCGGATCGTGATCTACAGTACATGAGACGACCTGACGTGGGAGCG 2413
Db 5075 ACGGCATCGATGGCGGATCGTGATCTACAGTACATGAGACGACCTGACGTGGGAGCG 5134
QY 2414 GGGGCCCTAGGATCGATTAAAGCTTCGCGGGGCTAGCACCGGT 2457
Db 5135 GGGGCCCTAGGATCGATTAAAGCTTCGCGGGGCTAGCACCGGT 5178
```

```
RESULT 12
US-10-190-435-13
; Sequence 13, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEYER, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 3531
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagPolmut_C
US-10-190-435-13
```

Query Match 96.8%; Score 2383.6; DB 14; Length 3531;  
Best Local Similarity 98.8%; Pred. No. 0;

```
Matches 2414; Conservative 0; Mismatches 24; Indels 6; Gaps 1;
QY 14 TGGCCGAGGCGCATGAGCCAGCCACAGCGCCAAACATCTCTGATGCGGAGCAACTTCA 73
Db 1088 TGGCCGAGGCGCATGAGCCAGCCACAGCGCCAAACATCTCTGATGCGGAGCAACTTCA 1147
QY 74 AGGGCCCCCAAGCGCATCATCAAGTGTCTCAACTCGCGCAAGGAGGGCCACATCGCCCGCA 133
Db 1148 AGGGCCCCCAAGCGCATCATCAAGTGTCTCAACTCGCGCAAGGAGGGCCACATCGCCCGCA 1207
QY 134 ACTGCGCGCCCCCGGCAAGAGGGCTGCTGGAAGTGGGCAAGGAGGGCCACAGATGA 193
Db 1208 ACTGCGCGCCCCCGGCAAGAGGGCTGCTGGAAGTGGGCAAGGAGGGCCACAGATGA 1267
QY 194 AGGACTGCAACGAGCGCGCAGGCCCAACTTCTTCCGCGAGGACCTGGCTTCCGCCAGGGCA 253
Db 1268 AGGACTGCAACGAGCGCGCAGGCCCAACTTCTTCCGCGAGGACCTGGCTTCCGCCAGGGCA 1327
QY 254 AGGCGCGGAGTTCCTCCGAGGAGCAACCGCGCCNACAGCCGCCACAGCGCGGAGCTGC 313
Db 1328 AGGCGCGGAGTTCCTCCGAGGAGCAACCGCGCCNACAGCCGCCACAGCGCGGAGCTGC 1387
QY 314 AGGTGCGCGCGACAAACCCCGCAGCGAGCGCGCGCGCGAGCGCCAGGCAACCTGAACT 373
Db 1388 AGGTGCGCGCGACAAACCCCGCAGCGAGCGCGCGCGCGAGCGCCAGGCAACCTGAACT 1447
QY 374 TCCCCAGATACCTCTGTGGGAGCGCCCTCTGGTGAGCATCAAGTGGGCGGCGGAGATCA 433
Db 1448 TCCCCAGATACCTCTGTGGGAGCGCCCTCTGGTGAGCATCAAGTGGGCGGCGGAGATCA 1507
QY 434 AGGAGGCCCTGTGACACCGCGCGCGAGCACACCGTGTCTGAGGAGATGAGCTTCCCG 493
Db 1508 AGGAGGCCCTGTGACACCGCGCGCGAGCACACCGTGTCTGAGGAGATGAGCTTCCCG 1567
QY 494 GCAAGTGAAGGCCCAAGATGATCGCGGCGCATCGCGGGCTTCAATCAAGGTGGCGCAGTAC 1627
Db 1568 GCAAGTGAAGGCCCAAGATGATCGCGGCGCATCGCGGGCTTCAATCAAGGTGGCGCAGTAC 1627
QY 554 ACCAGATCTGATCGAGATCTCGGCAAGAGGCCCATCGGACCGCTGTGTGATCGGCCCA 613
Db 1628 ACCAGATCTGATCGAGATCTCGGCAAGAGGCCCATCGGACCGCTGTGTGATCGGCCCA 1687
QY 614 CCCCCGTGAACATCATTCGGCCCGCAACATGCTGACCCAGCTGGGGCTGCAACCTTGAAC 673
Db 1688 CCCCCGTGAACATCATTCGGCCCGCAACATGCTGACCCAGCTGGGGCTGCAACCTTGAAC 1747
QY 674 CCATCAGCCCCATCGAGACCGCTGGAAGCTGAAGCCCGGCGCATGACGCGCCCAAGG 733
Db 1748 CCATCAGCCCCATCGAGACCGCTGGAAGCTGAAGCCCGGCGCATGACGCGCCCAAGG 1807
QY 734 TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTCGGAGGAGA 793
Db 1808 TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTCGGAGGAGA 1867
QY 794 TGAAGAGGAGGCGCAAGATCAACAGATCGCGCGCGGAGAACCCCTACACACCCCGTGT 853
Db 1868 TGAAGAGGAGGCGCAAGATCAACAGATCGCGCGCGGAGAACCCCTACACACCCCGTGT 1927
QY 854 TCGCCCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db 1928 TCGCCCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 1987
QY 914 ACAAGGCGACCCAGGACTTCTGGGAGGTGAGTGGGCGATCCGCCACCCCGCGGCGTGA 973
Db 1988 ACAAGGCGACCCAGGACTTCTGGGAGGTGAGTGGGCGATCCGCCACCCCGCGGCGTGA 2047
QY 974 AGAAGAGAGAGAGCGTGACCGTCTCTGGAGGTGGGCGAGCGCTACTTCAAGCTGCCCTGG 1033
Db 2048 AGAAGAGAGAGAGCGTGACCGTCTCTGGAGGTGGGCGAGCGCTACTTCAAGCTGCCCTGG 2107
QY 1034 ACGAGGACTTCCGCAAGTACACCGCTTCAATCCCGAGCATCAACAGAGAGAGAGAGAGAG 1093
Db 2108 ACGAGGACTTCCGCAAGTACACCGCTTCAATCCCGAGCATCAACAGAGAGAGAGAGAGAG 2167
```





```

; GENERAL INFORMATION:
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: P18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 3537
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagPolmutina_C
; US-10-190-435-15

Query Match          96.7%  Score 2382;  DB 14;  Length 3537;
Best Local Similarity 98.7%  Pred. No. 0;
Matches 2413;  Conservative 0;  Mismatches 25;  Indels 6;  Gaps 1;

Qy 14  TGGCGAGGCGATGAGCGAGGCGACCGAGCGCAACATCTGATGAGCGCGAGCAACTTCA 73
Db 1094 TGGCGAGGCGATGAGCGAGGCGCAACCGAGCGTGTATGATGAGAGAGCAACTTTAAAA 1153

Qy 74  AGGCGCCCAAGCGATCATCAAGTGTCTCACTGCGGCAAGGAGGCGCACATCGCCGCA 133
Db 1154 AGGCGCCCAAGCGATCATCAAGTGTCTCACTGCGGCAAGGAGGCGCACATCGCCGCA 1213

Qy 134 ACTGCGCGCGCCCGCGAGAGAGGGGTCTGTGAAGTGTGCGCAAGAGGCGCACCGATGA 193
Db 1214 ACTGCGCGCGCCCGCGAGAGAGGGGTCTGTGAAGTGTGCGCAAGAGGCGCACCGATGA 1273

Qy 194 AGGACTGACCGAGCGCGAGGCGCAACTTCTTCGCGAGGAGCTGTGCGCTTCCCGAGGSCA 253
Db 1274 AGGACTGACCGAGCGCGAGGCGCAACTTCTTCGCGAGGAGCTGTGCGCTTCCCGAGGSCA 1333

Qy 254 AGGCGCGAGGTTCCCGAGGAGAGCAACCGGCGCAAGCGCCCAAGCGCGCGAGCTGC 313
Db 1334 AGGCGCGAGGTTCCCGAGGAGAGCAACCGGCGCAAGCGCCCAAGCGCGCGAGCTGC 1393

Qy 314 AGGTGCGCGCGCAACACCCCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373
Db 1394 AGGTGCGCGCGCAACACCCCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1453

Qy 374 TCCCGCAGATCACTCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433
Db 1454 TCCCGCAGATCACTCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1513

Qy 434 AGGAGGCGCTGTGCGACACCGGCGCGCGAGACACCGCTGTGAGGAGATGAGCTGCGCG 493
Db 1514 AGGAGGCGCTGTGCGCGCACCGGCGCGCGAGCACCGCTGTGAGGAGATGAGCTGCGCG 1573

Qy 494 GCAAGTGGAGAGCCCAAGATGATCGCGCGATCGCGCGCTTCAAGGTGCGCGCGAGTACG 553
Db 1574 GCAAGTGGAGAGCCCAAGATGATCGCGCGATCGCGCGCTTCAAGGTGCGCGCGAGTACG 1633

Qy 554 ACCAGATCTCTGATCGAGATCTGCGGCAAGAGCGCATCGCGACCGCTGTGATCGCGCGCA 613
Db 1634 ACCAGATCTCTGATCGAGATCTGCGGCAAGAGCGCATCGCGACCGCTGTGATCGCGCGCA 1693

Qy 614 CCGCGGTGAACATCATCGCGCGCAACATGTGACCGCGCTGTGCGCTGTGCGCTGTGCGCT 673
Db 1694 CCGCGGTGAACATCATCGCGCGCAACATGTGACCGCGCTGTGCGCTGTGCGCTGTGCGCT 1753

Qy 674 CCATCAGCCCGATCGAGACCGTGTGCGCGTGAAGTGAAGCGCGCGCGCGCGCGCGCGCG 733
Db 1754 CCATCAGCCCGATCGAGACCGTGTGCGCGTGAAGTGAAGCGCGCGCGCGCGCGCGCGCG 1813

```

```

734 TGAAGCAGTGGCCCTCTGACCGAGGAGAGATCAAGGCGCTTACCGCCATCTTGCAGAGAGA 793
1814 TGAAGCAGTGGCCCTCTGACCGAGGAGAGATCAAGGCGCTTACCGCCATCTTGCAGAGAGA 1873

Qy 794 TGGAGAGAGAGGCGAGATCAACCAAGATCGGCGCGGAGAAACCCCTACAAACACCCCGTGT 853
1874 TGGAGAGAGAGGCGAGATCAACCAAGATCGGCGCGGAGAAACCCCTACAAACACCCCGTGT 1933

Qy 854 TGGCCATCAAGAGAGAGAGACCAACCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGTGA 913
1934 TGGCCATCAAGAGAGAGAGACCAACCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGTGA 1993

Qy 914 ACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCGACCCCGCGCGCTGA 973
1994 ACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCGACCCCGCGCGCTGA 2053

Qy 974 AGAAGAGAGAGCGGTGACCGGTGCTGGAGCGTGGGCGAGCGCTTCTAGCGTGGCGCTGG 1033
2054 AGAAGAGAGAGCGGTGACCGGTGCTGGAGCGTGGGCGAGCGCTTCTAGCGTGGCGCTGG 2113

Qy 1034 ACGAGAGCTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACAAACAGAGACCCCG 1093
2114 ACGAGAGCTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACAAACAGAGACCCCG 2173

Qy 1094 GCATCCGCTACCAAGTACCAAGTGTGCGCGAGGCGTGGAGGCGAGCGCGCGAGCTTTC 1153
2174 GCATCCGCTACCAAGTACCAAGTGTGCGCGAGGCGTGGAGGCGAGCGCGCGAGCTTTC 2233

Qy 1154 AGAGCAGCATGACCAAGATCTCTGGAGCGCTTCCGCGCGCGCAACCCCGAGATCTGTATCT 1213
2234 AGAGCAGCATGACCAAGATCTCTGGAGCGCTTCCGCGCGCGCAACCCCGAGATCTGTATCT 2293

Qy 1214 ACCAGCGCGCTGTGAGTGGGCGTGGAGTGGCGAGTGGCGAGTGGCGAGTGGCGAGTGG 1273
2294 ACCAGCGCGCTGTGAGTGGGCGTGGAGTGGCGAGTGGCGAGTGGCGAGTGGCGAGTGG 2353

Qy 1274 AGAGCTGCGCAAGCACTGTGCGTGGGCGTTCACCAACCCCGCAAGAGCAACAGCA 1333
2354 AGAGCTGCGCAAGCACTGTGCGTGGGCGTTCACCAACCCCGCAAGAGCAACAGCA 2413

Qy 1334 AGAGCGCGCTTCTGTGTGGTGGGCTACGAGTGGGCTACCGCGCAAGTGGAGTGGGCG 1393
2414 AGAGCGCGCTTCTGTGTGGTGGGCTACCGCGCAAGTGGAGTGGGCGTGGGCG 2467

Qy 1394 CCATCGAGCTGCCCGAGAGAGAGTGGACCGTGGAGAGTGGACCGTGGAGAGTGGAG 1453
2468 CCATCGAGCTGCCCGAGAGAGAGTGGACCGTGGAGAGTGGACCGTGGAGAGTGGAG 2527

Qy 1454 AGCTGAATGGGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1513
2528 AGCTGAATGGGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2587

Qy 1514 TGGCGCGCGCGAGCGCTGACCGCATCTGTCGCGCTTACCGAGAGGCGCGAGTGGAGC 1573
2588 TGGCGCGCGCGAGCGCTGACCGCATCTGTCGCGCTTACCGAGAGGCGCGAGTGGAGC 2647

Qy 1574 TGGCGCGAGAACCGCGAGATCTCTGCGCGAGCGCGTGGAGCGCGTGTACTAGACCCCA 1633
2648 TGGCGCGAGAACCGCGAGATCTCTGCGCGAGCGCGTGGAGCGCGTGTACTAGACCCCA 2707

Qy 1634 AGAAGCTGTGGCGCGAGATCTGAGAGAGGCGCGAGCGAGTGGAGTGGAGTGGAGTGG 1693
2708 AGAAGCTGTGGCGCGAGATCTGAGAGAGGCGCGAGCGAGTGGAGTGGAGTGGAGTGG 2767

Qy 1694 AGAGCGCTTCAAGAACCTGGAAGACCGGCAAGTGGCGCAAGTGGCGCAAGTGGCGCA 1753
2768 AGAGCGCTTCAAGAACCTGGAAGACCGGCAAGTGGCGCAAGTGGCGCAAGTGGCGCA 2827

Qy 1754 AGAAGTGAAGAGCTGAGCGAGGCGCGTGGAGAGATGGCGATGGAGAGATGGAGATGG 1813
2828 AGAAGTGAAGAGCTGAGCGAGGCGCGTGGAGAGATGGCGATGGAGAGATGGAGATGG 2887

Qy 1814 GGGGCAAGACCCCGCAAGTTCGCGCTTCCCGCTTCCCGCTTCCCGCTTCCCGCTTCC 1873

```

```

Db      2888  GGCGCAAGACCCCAAGTTCCGCTGCTCCCATCCAGAGAGAGACCTGGAGACCTTGTGGA 2947
Qy      1874  CCGACTACTGGGAGGCCACTGTGATCCCGAGTGGGAGTTGTGAACACCCGCCCTCTGG 1933
Db      2948  CCGACTACTGGGAGGCCACTGTGATCCCGAGTGGGAGTTGTGAACACCCGCCCTCTGG 3007
Qy      1934  TGAAGTGTGTGTTACCAAGCTGGAGAGAGAGCCCATCATCGGGCCGAGACCTTCTACGTGG 1993
Db      3008  TGAAGTGTGTGTTACCAAGCTGGAGAGAGAGCCCATCATCGGGCCGAGACCTTCTACGTGG 3067
Qy      1994  ACGGCCCGCCCAACCCGAGACCAAGATCGGAAGCGCGCTGTACGTACCGACCGGGGCC 2053
Db      3068  ACGGCCCGCCCAACCCGAGACCAAGATCGGAAGCGCGCTGTACGTACCGACCGGGGCC 3127
Qy      2054  GGCAAGATCGTGTGAGCTTGACCGAGACCAACCAAGAGAGAGAGAGAGAGAGAGAGAG 2113
Db      3128  GGCAAGATCGTGTGAGCTTGACCGAGACCAACCAAGAGAGAGAGAGAGAGAGAGAGAG 3187
Qy      2114  AGCTGCCCTGAGGACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2173
Db      3188  AGCTGCCCTGAGGACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3247
Qy      2174  TGGGATCATCCAGGCCCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2233
Db      3248  TGGGATCATCCAGGCCCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3307
Qy      2234  AGCAGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2293
Db      3308  AGCAGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3367
Qy      2294  GGGCAACAGAGAGATCGCAAGCTGTGTGAGCAAGGGGATCCGCAAGTGTGTTCCTGG 2353
Db      3368  GGGCAACAGAGAGATCGCAAGCTGTGTGAGCAAGGGGATCCGCAAGTGTGTTCCTGG 3427
Qy      2354  ACGGCATCATGCGGCGATCGTATCTACAGTGTGTGAGCAAGGGGATCCGCAAGTGTGT 2413
Db      3428  ACGGCATCATGCGGCGATCGTATCTACAGTGTGTGAGCAAGGGGATCCGCAAGTGTGT 3487
Qy      2414  GGGCCCTAGGATCGATTAAAGCTTCCGGGGCTAGCACCGGT 2457
Db      3488  GGGCCCTAGGATCGATTAAAGCTTCCGGGGCTAGCACCGGT 3531

```

RESULT 15

```

US-10-190-435-12
; Sequence 12, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrellita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GagCompPolmutInaTatRevNef_C
US-10-190-435-12

```

Query Match 96.7%; Score 2381; DB 14; Length 5145;  
 Best Local Similarity 99.5%; Pred. No. 0;  
 Matches 2400; Conservative 0; Mismatches 5; Indels 6; Gaps 1;

```

Qy      14  TGGCCGAGGCCATGAGCCAGGCCACAGCGCCCAACATCTCTGATGCGAGCGAGCAACTTCA 73
Db      1487  TGGCCGAGGCCATGAGCCAGGCCACAGCGCCCAACATCTCTGATGCGAGCGAGCAACTTCA 1546
Qy      74  AGGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGAGAGGGCCACATCGCCCGCA 133
Db      1547  AGGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGAGAGGGCCACATCGCCCGCA 1606
Qy      134  ACTGCGCGGCCCGCCGCAAGAGAGGGCTGTGGAAGTGGCGCAAGAGAGGGCCACCATGTA 193
Db      1607  ACTGCGCGGCCCGCCGCAAGAGAGGGCTGTGGAAGTGGCGCAAGAGAGGGCCACCATGTA 1666
Qy      194  AGGACTGACACGAGCGCCAGCCAACTTCTTCGCGAGGACCTTGGCTTCCCGCAGGGCA 253
Db      1667  AGGACTGACACGAGCGCCAGCCAACTTCTTCGCGAGGACCTTGGCTTCCCGCAGGGCA 1726
Qy      254  AGGCGCCGAGCTTCCCGAGCGAGCAGAACCGCCCAACAGCCCAACAGCCCAACAGCCG 313
Db      1727  AGGCGCCGAGCTTCCCGAGCGAGCAGAACCGCCCAACAGCCCAACAGCCCAACAGCCG 1786
Qy      314  AGGTGCGCGCGCAACAAACCCCGCAGCGAGCCCGCGCGAGCCCGAGCGCCAGGGCACCT 373
Db      1787  AGGTGCGCGCGCAACAAACCCCGCAGCGAGCCCGCGCGAGCGCCAGGGCACCTGAACT 1846
Qy      374  TCCCGCCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGCGGCGCAGATCA 433
Db      1847  TCCCGCCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGCGGCGCAGATCA 1906
Qy      434  AGGAGGCCCTGCTGCAACACCGCGGCCGACACACCTGTGTGGAGGAGATGAGCTGCCCG 493
Db      1907  AGGAGGCCCTGCTGCGCCACACCGCGGCCGACACACCTGTGTGGAGGAGATGAGCTGCC 1966
Qy      494  GCAAGTGAAGCCCAAGATGATCGCGGGCATCGCGGGCTTCAACAAGTGGCGGAGTACG 553
Db      1967  GCAAGTGAAGCCCAAGATGATCGCGGGCATCGCGGGCTTCAACAAGTGGCGGAGTACG 2026
Qy      554  ACCAGATCCTGATCGAGATCTGGCGCAAGAGGCCATCGGCACCGTCTGATCGGCCCA 613
Db      2027  ACCAGATCCTGATCGAGATCTGGCGCAAGAGGCCATCGGCACCGTCTGATCGGCCCA 2086
Qy      614  CCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTC 673
Db      2087  CCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTC 2146
Qy      674  CCATCAGCCCATCGAGACCGTGCCTGAGCTGAAGCTGAAGCCCGCATGGAGCGGCCAAG 733
Db      2147  CCATCAGCCCATCGAGACCGTGCCTGAGCTGAAGCTGAAGCCCGCATGGAGCGGCCAAG 2206
Qy      734  TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTCGGAGGAGA 793
Db      2207  TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTCGGAGGAGA 2266
Qy      794  TGGAGAGAGGGCAAGATCAACAGATCGGCCCGCCGAGAACCCCTACACACCCCGTGT 853
Db      2267  TGGAGAGAGGGCAAGATCAACAGATCGGCCCGCCGAGAACCCCTACACACCCCGTGT 2326
Qy      854  TGGCCATCAAGAAGAAGAGAGACCAAGTGGCGCAAGCTGGTGGAGTTCGCGGAGCTGA 913
Db      2327  TGGCCATCAAGAAGAAGAGAGACCAAGTGGCGCAAGCTGGTGGAGTTCGCGGAGCTGA 2386
Qy      914  ACAAGCGCACCCAGGACTTCTGGAGAGTGCAGTGGGCTATCCCGCATCCCGCGGCTGA 973
Db      2387  ACAAGCGCACCCAGGACTTCTGGAGAGTGCAGTGGGCTATCCCGCATCCCGCGGCTGA 2446
Qy      974  AGAAGAAGAGAGCTGACCGTGTGAGAGCTGGCGGACGCTACTTTCAGGGTCCCGCTGG 1033
Db      2447  AGAAGAAGAGAGCTGACCGTGTGAGAGCTGGCGGACGCTACTTTCAGGGTCCCGCTGG 2506
Qy      1034  ACGAGGACTTCCGCAAGTACACCGCTTCACATCCCGCAGCATCAACAGAGAGACCCCG 1093
Db      2507  ACGAGGACTTCCGCAAGTACACCGCTTCACATCCCGCAGCATCAACAGAGAGACCCCG 2566

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 06:34:42 ; Search time 4148.08 Seconds  
(without alignments)  
17774.420 Million cell updates/sec

Title: US-09-610-313-30  
Perfect score: 2469  
Sequence: 1 gtcagccaccatgcgcga.....ggctagcaccggtagaattc 2469

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estnu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hcc.\*  
9: gb\_estl.\*  
10: gb\_est2.\*  
11: gb\_hcc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_mam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rpd.\*  
26: em\_gss\_phg.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*  
29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	90.8	3.7	2598	11	AY103647 Zea mays
2	85.6	3.5	951	12	BM321451
3	85	3.4	869	14	CK159167
4	80.2	3.2	1132	12	BM320864

5	79.6	3.2	1165	12	BM320900
6	79	3.2	867	12	BM321430
7	76.6	3.1	1550	12	BM321022
8	75.6	3.1	671	13	CA093222
9	75.4	3.1	2299	11	AY106831
10	75	3.0	862	12	BM321023
11	74.6	3.0	914	28	BZ568300
12	74.4	3.0	3134	11	AY109500
13	73.8	3.0	889	14	CK159613
14	73.6	3.0	757	29	CG578788
15	73.2	3.0	853	12	BM321333
16	72.8	2.9	566	12	BM567428
17	72.6	2.9	788	14	CB643171
18	72.4	2.9	753	29	CC675888
19	72.4	2.9	801	12	BJ375394
20	72.2	2.9	688	14	CB648640
21	72.2	2.9	764	14	CB651670
22	72.2	2.9	766	14	CB642928
23	72.2	2.9	809	14	CB641357
24	72.2	2.9	841	14	CB651502
25	72.2	2.9	851	29	CG260623
26	72.2	2.9	860	29	CG691066
27	71.8	2.9	1933	11	AY325173
28	71.8	2.9	5622	29	AY400359
29	71.2	2.9	646	12	BM645449
30	71.2	2.9	724	12	BM584131
31	71	2.9	725	14	CB630209
32	71	2.9	745	14	CB641936
33	71	2.9	766	14	CB629976
34	71	2.9	787	14	CB658685
35	71	2.9	792	14	CB618305
36	71	2.9	810	14	CB618374
37	71	2.9	823	14	CB627740
38	71	2.9	865	14	CB628660
39	71	2.9	925	29	CNS0091P
40	70.6	2.9	703	14	CB675843
41	70.6	2.9	710	14	CB676048
42	70.6	2.9	712	14	CB647546
43	70.6	2.9	714	14	CB651194
44	70.6	2.9	715	14	CB646590
45	70.6	2.9	715	14	CB669855

## ALIGNMENTS

RESULT 1	AY103647	2598 bp	mRNA	linear	HTC 16-OCT-2002
LOCUS	AY103647	Zea mays	PC0142084	mRNA sequence.	
DEFINITION	AY103647	Zea mays	PC0142084	mRNA sequence.	
ACCESSION	AY103647	HTC.			
VERSION	AY103647.1	GI:21206725			
KEYWORDS	HTC.				
SOURCE	Zea mays				
ORGANISM	Zea mays				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.				
AUTHORS	Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.				
TITLE	Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes				
JOURNAL	Unpublished (2002)				
REFERENCE	2 (bases 1 to 2598)				
AUTHORS	Coe, E.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA				
COMMENT	If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the				



maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

Location/Qualifiers

1..2598

/organism="Zea mays"

/mol\_type="mRNA"

/db\_xref="MaizeDB:638378"

/db\_xref="taxon:4577"

/clone\_lib="Maize Mapping Project/DuPont Cornsensus Library"

/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

# ORIGIN

```

Query Match      3.7%; Score 90.8; DB 11; Length 2598;
Best Local Similarity 43.3%; Pred. No. 0.21;
Matches 719; Conservative 0; Mismatches 909; Indels 33; Gaps 5;

170 GCGGCAAGGAGGCGACCGAGTGAAGGAGTGCACCGAGCGCCAGGCGCAACTTCTTCGCG 229
351 GCAGCGAGTGATCTCGATCGATCGCCACCGCGCGGACGAGCGCGGCTTCCAGCGTCA 410
230 AGGACCTGGCGCTTCCCGGCGGCGGCGGCGGAGTTCGCCAGCGAGCAGAGACCGCGCA 289
411 GCTTCGGCGGCGACCTCCACCGCGGCGGCGGCGGCGGCGGCTTCCAGGTCGAGTGC 470
290 ACAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 349
471 TCAGCGGATCTCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 530
350 CCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 409
531 AGGTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 590
410 GCATCAAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 469
591 TCCGCTTCGAGATCTCCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 650
470 TGTGGAGGAGTGAAGTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 529
651 TGCGCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 710
530 GTTTCATCAAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 589
711 GCCTCATCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 767
590 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 649
768 ACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 827
650 AGCTGGGCGTGCACCTGAACTTCCCATATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 709
828 AGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 887
710 AGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 769
888 TGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 947
770 CCCTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 829
948 TCATGAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1007
830 AGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 899
1008 GTTCATCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1067
890 AGCTGGTGGAGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 949

```

## FEATURES

source

```

1068 AGGCCAAGAAGTTGAACAGAGTGGACCCCGC-----TGCTGAAGCCCAAGCAGG 1115
950 GCATCCCCCACCACCGCGCGCTGAAGAAGAAGAAGAGCGTACCCTGTGTGACGTGGCGG 1009
1116 ACAGGTACGCGCTCCGACGTCGCGCAGTGTGCGGCGCCAGATCAGAGTCAATCCGCG 1175
1010 AGCCTACTTACGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1069
1176 CCGCCACCAAGTCCATCGAGCGCGAGGTCAATCCGCGTGAACGACCAACCCGCTCATGAG 1235
1070 CCAGCATCAACACGAGAGACCCCGCGCATCCGCTACCCAGTACCAACGCTGTGCCCGAGG 1129
1236 TCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1295
1130 GGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1189
1296 TGGACAAACGCGCGCTCGCGCATCGCCAAACATCGGCAAGCTCATGTTGCGCGAGTTTCCG 1355
1190 CCGCAACCCCGGAGATCGTGATCTACAGTACATGACGACCTGTACGTGGGCGAGCGA-- 1247
1356 AGTCTGTCAAGAGTCTTACAAACGCGCTCACTCCAACTTGGCGGCGGCGGCGGCGGCGG 1415
1248 ----CTTGGAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1303
1416 CCAGCTTGGAGTACGCGCTTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1475
1304 GCTGGGCTTCAACACCCCGGAGAGCAAGCAAGCAAGAGGAGCGGCGGCGGCGGCGGCGGCG 1363
1476 TCCAGTACCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1535
1364 GCTACGAGCTGCACCCCGGAGTGGACCGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1423
1536 AGTGAACTCCCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1595
1424 GCTGGACGCTGAACGACATCCAGAGCGTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1483
1596 AGTCTATGTCTCCACCTACATGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1655
1484 ACCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1543
1656 AGGAGAACATCAAGCGCTCGGTGAAGAACACCGGTGACCCAGGTGGCGGCGGCGGCGGCG 1715
1544 ACATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1594
1716 CCATGAACCCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1775
1595 AGATCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1654
1776 CCATGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1835
1655 AGATCCAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1714
1836 TGATGAGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1892
1715 ACCTGAAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1774
1893 ACGGAGACCCCTCGGTGTTCTCCAGATCACCAGGTTCCGAGGAGAGTCCGCGCGGTGC 1952
1775 TGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1815
1953 TGCCCCGAGGAGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1993

```

## RESULT 2

BM321451

LOCUS

DEFINITION

rockefeller.0.1211 Mastigamoeba balamuthi lambda ZAP II Library

Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC

3.3.1.1), mRNA sequence.

BM321451

EST.

KEYWORDS

SOURCE

BM321451 951 bp mRNA linear EST 03-JAN-2002  
 rockefeller.0.1211 Mastigamoeba balamuthi lambda ZAP II Library  
 Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC  
 3.3.1.1), mRNA sequence.  
 BM321451  
 EST.  
 GI:18055857  
 Mastigamoeba balamuthi

ORGANISM Mastigamoeba balamuthi  
 Eukaryota; Pelobiontidae; Mastigamoebidae; Mastigamoeba.  
 REFERENCE 1 (bases 1 to 951)  
 AUTHORS Baptiste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Durfee, L., Gaasterland, T., Lopez, P., Muller, M. and Philippe, H.  
 TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
 MEDLINE 21819461  
 PUBMED 11830664  
 COMMENT Contact: Muller Miklos  
 Laboratory of Biochemical Parasitology  
 The Rockefeller University  
 1230 York Avenue, New York, NY 10021, USA  
 Email: mmuller@rockefeller.edu  
 Insert Length: 951 Std Error: 0.00  
 POLYA-No.  
 FEATURES Location/Qualifiers  
 source 1..951  
 /organism="Mastigamoeba balamuthi"  
 /mol\_type="mRNA"  
 /strain="ATCC 30984"  
 /db\_xref="taxon:108607"  
 /clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
 /note="syn: Phreatamoeba balamuthi"  
 ORIGIN  
 Query Match 3.5%; Score 85.6; DB 12; Length 951;  
 Best Local Similarity 46.0%; Pred. No. 0.73;  
 Matches 323; Conservative 0; Mismatches 376; Indels 3; Gaps 1;  
 QY 1734 CGCCAGATGGCAGCCGCCACACCAAGCAGTGAAGCAGTACCGAGGCGGTGCAGAA 1793  
 Db 219 CGCCAGATGGCAGTGGTGTCTGTCGACATCTTCGACGAGGACACGCGCGCGC 278  
 QY 1794 GATGCCATGAGAGCATCTGATCTGGGGCAAGACCCCAAGTTCGCTGCCATCCA 1853  
 Db 279 CATGCGCAGCGCGCTCTGGTCTTCTGCTGAGGGCGAGACTTCAGGAGTACTG 338  
 QY 1854 GAAGAGACCTGGAGACCTGGTGACCGATATGCGAGGCCACCTGGATGCCCGAGTG 1913  
 Db 339 GGAGTGACCTGGAAGGCGCTGTCTCGGCCCTACCGAGGCGCTCAGATCATCTGCGA 398  
 QY 1914 GGAGTTCGTGACACCCCGCCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCAT 1973  
 Db 399 CGACGGCGGTGACGCGACTCTGATGATCCCAAGGGGTTCCGCGCGAGGACACCCCAA 458  
 QY 1974 CATCGGCCCGAGACCTTCTACGTGGACGCGCGCCCAACCGGAGCCAGATCGGCAA 2033  
 Db 459 GCTGCTGGAGACGACGAGGCGCTCGAGAGGTGCGCTGCTCAACACGTGCTCAAGCA 518  
 QY 2034 GCGCGGCTACGTGACCGACCGGGCGCGCAGAGATCGTGAGCTGACCGGAGACCCAA 2093  
 Db 519 GGTCCAGAAGGAGCAGCGCGCTTCTGCGACAAGATCTCTCCCGAGATCCGCGGTGAG 578  
 QY 2094 CCAGAGACCGAGCTGAGGCGCATCCAGTGGCCCTGACGACGAGCGGACCGAGTGA 2153  
 Db 579 CGAGGAGACGAGACTGGCGTATGAGCTGTACAGCTGACCGCGAGCGCAAGTGTCT 638  
 QY 2154 CATCTGTGACCGACAGCAGTACGCGCTGGGATCATCCAGGCGCCAGCCGACAGAGCGA 2213  
 Db 639 GTTCCGCGCGTCAAGGTCAAGCAGC---TCTNTCAACCAAGCAAGTTCGACACATCTA 695  
 QY 2214 GAGCAGCTGTGTGAACCAAGATCATCGAGCAGTGTATCAAGAGAGAGAGTGTACTTAG 2273  
 Db 696 CGGCTGCGCGCCACTCGCTCATCGACGGCATCAAGCGCGGCGACCGACGTGATGCTCG 755  
 QY 2274 CTGGGTGCCCGCCCAAGAGGCGATCGCGGCAACCGAGCAGATCGACAGCTGGTGAGCAA 2333  
 Db 756 CAAGGTGCGCGTCTGCGGGGCTACGCGAGAGTGGGCAAGGGCTGCGCGGAGTCTGCTGG 815  
 QY 2334 GGGCATTCGCAAGGTGTCTTCTTCTGGAACGGATCATGCGCGGCATCGTGATCTACAGTA 2393

Db 816 CGCCAGGCGCTGCGCGCTCATCTGTCACGAGATCGACCCCATCTGCGGCTGCGAGGCTC 875  
 QY 2394 CATGACGACCTGTACGTGGCAGCGGCGGCGCTTAGGATCGA 2435  
 Db 876 GATGCGCGCTTCGAGGTCAACACGCTCAGGCGGGGCTCGA 917  
 RESULT 3  
 LOCUS CK159167/c  
 DEFINITION FGAS040564 Triticum aestivum FGAS: TaLts5 Triticum aestivum cDNA,  
 mRNA sequence.  
 ACCESSION CK159167  
 VERSION CK159167.1 GI:38985053  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.  
 REFERENCE 1 (bases 1 to 869)  
 AUTHORS Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D., Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A., Links, V.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D., Pennik, C., Roach, J.L. and Sarhan, F.  
 TITLE Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Wm L Crosby  
 Bioinformatics  
 University of Saskatchewan, Department of Computer Science  
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
 Saskatchewan, S7N 5A9, Canada  
 Tel: 306 966 1769  
 Fax: 306 966 2033  
 Email: fgas.ests@usask.ca  
 This sequence is the direct result of the base calling software  
 Phred (default parameters). It is the raw base calls. To aid in the  
 identification of the high quality insert the software Lucy  
 (default parameters) has been run on this sequence. Lucy identified  
 the region [128,636].  
 Plate: TaLts57 row: N column: 23.  
 FEATURES Location/Qualifiers  
 source 1..869  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Wheat line PI 178383"  
 /db\_xref="taxon:4565"  
 /lab\_host="DH5 alpha"  
 /clone\_lib="Triticum aestivum FGAS: TaLts5"  
 /note="Organ: Crown; Vector: pGEM-T; SSH (suppression  
 subtractive hybridization) cDNA library from genotype  
 PI178383 cold hardened at 2 C for 21 days and 49 days  
 (equal amount of cDNA pooled together before subtraction,  
 tester) and subtracted against genotype Norstar cold  
 hardened at 2 C for 1 day (24 H) (driver). Modified Smart  
 cDNA (Clontech) priming and non-directional cloning"  
 ORIGIN  
 Query Match 3.4%; Score 85; DB 14; Length 869;  
 Best Local Similarity 45.4%; Pred. No. 0.85;  
 Matches 304; Conservative 0; Mismatches 365; Indels 0; Gaps 0;  
 QY 580 AGAAGCGCATCGGACCGTGTGATCGGCCCGCCCGCTGATCATCGCGCGCAC 639  
 Db 827 ATGGCGCGCGCGCACACGAGGTATCCACACGAGGAGACACACACACACAC 768  
 QY 640 ATGCTGACCCAGTGGGCTGCACCTTGAACTTCCCATCAGCCCATCGAGACCTG 699  
 Db 767 CACAACGACGACAAACAGCGACACACACACACACACACACACACACACAC 708  
 QY 700 GTGAAGTGAAGCGCGCATGAGCGGCCCGCCCGAGGTGAGCAGTGGCCCTGAC 759



```

KEYWORDS      EST.
SOURCE         Mastigamoeba balamuthi
ORGANISM       Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE      1 (bases 1 to 1165)
AUTHORS        Baptiste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
                Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
                Philippe,H.
TITLE          The analysis of 100 genes supports the grouping of three highly
                divergent amoebae: Dictyostellium, Entamoeba, and Mastigamoeba
JOURNAL        Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE        21819461
PUBMED        11830664
COMMENT        Contact: Muller Miklos
                Laboratory of Biochemical Parasitology
                The Rockefeller University
                1230 York Avenue, New York, NY 10021, USA
                Email: mmuller@rockvax.rockefeller.edu
                Insert Length: 1165 Std Error: 0.00
                POLYA=No.

FEATURES             Location/Qualifiers
     source           1..1165
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"

ORIGIN
Query Match      3.2%; Score 79.6; DB 12; Length 1165;
Best Local Similarity 45.1%; Pred. No. 4;
Matches 428; Conservative 0; Mismatches 506; Indels 16; Gaps 3;

QY 90 CATCAAGTCTTAACTCGGAGAGGGCCACATGCCCCGCACTGCGCGCCCGG 149
DB 46 COTCAAGAACAAAGCGGTACTTCAAGCGCTTCCAGACCCAGTTCGCGCCCGGAGGG 105
QY 150 CAAGAGGCGTCTGGAAGTGGCGCAAGAGGGCCACAGATGAAGGACTCCACCGAGCG 209
DB 106 CAAGAGGACTACCGCGCGCCGCACTGATCCAGACAAAGAACAGTACACAGCCC 165
QY 210 CAGGCCCAACTTCTTCGCGAGGACTTGGCTTCCCGAGGGCAAGGCCCGCGAGTTCCC 269
DB 166 CAAGTACCGCTTCGTC-----GTCCGCTTCAACCAAGGAGACATCGTCTCCAGATCGC 219
QY 270 CAGGAGCAGAACCCGCGCAACAGCCACAGCGCGAGCTGCGAGTGGCGCGCAAA 329
DB 220 CTACGCCAAGATCGACGGGACCAATCTCTGCGCGCCCTACTCGACAGATCAACCG 279
QY 330 CCCCAGCAGCGCGCGCGCGAGCGCCAGGCACTTGAATTTCCCGCAGATCACCGT 389
DB 280 CTTGCGCTCAAGTCTGCGCTGACCAATACGCGCGCCCTACGCGACTGCGCTGCTGCT 339
QY 390 GTGCGAGCGCCCTCTGTGAGCATCAAGTGGCGCGCCAGATCAAGGAGGCCCTGCTGGA 449
DB 340 GCGCCGCGCTGTGTGAAGAGTCAAACCTCTACTTCAAGTACGAGGCTGCAAGAGGT 399
QY 450 CACCGCGCGCGAGCACCGTCTCGAGGAGATGAGCTGCGCGCGAGTGGAGGCCCAA 509
DB 400 CAACGCGGAGGACTCAACGTCGAGGAGTTCGACGACGCGCGCGCGCGCTTCAAGGCC-- 457
QY 510 GATGATCGCGCGCATCGCGGCTTCATCAAGGTGGCGCCAGTACGACACAGATCTGATCGA 569
DB 458 --TGCTCGAGTCTGGCTGTTCGCGACCTCGACTGCGCGCGCGGTTCGCGCCCTCAA 515
QY 570 GATCTCGCGCAGAGAGCCATCGGACCGTGTGATGCGCGCCACCCCGGTGAACATCAT 629
DB 516 GGGCATGTCTCGAGCGCGCGCTCAACGTCGAGGAGTTCGCGCGCGCGCTTCAAGGCC-- 575
QY 630 CGGCGCGCAACATGCTGACCCAGTGGGTGACCCCTGAACTTCCCGCATCAGCGCCCATCGA 689
DB 576 CGGCGACAGAGAGGAGCTCAACGCGCGGTCTCTCCGCAAGTACATCTTTCGCGGCGCAGT 635

RESULT 6
LOCUS      BM321430
DEFINITION Mastigamoeba balamuthi cDNA similar to ribosomal protein S4, mRNA
ACCESSION  BM321430
VERSION     BM321430.1 GI:18055836
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
ORGANISM    Mastigamoeba balamuthi
REFERENCE    1 (bases 1 to 867)
AUTHORS      Baptiste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
                Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
                Philippe,H.
TITLE        The analysis of 100 genes supports the grouping of three highly
                divergent amoebae: Dictyostellium, Entamoeba, and Mastigamoeba
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE      21819461
PUBMED       11830664
COMMENT      Contact: Muller Miklos
                Laboratory of Biochemical Parasitology
                The Rockefeller University
                1230 York Avenue, New York, NY 10021, USA
                Email: mmuller@rockvax.rockefeller.edu
                Insert Length: 867 Std Error: 0.00
                POLYA=No.

FEATURES             Location/Qualifiers
     source           1..867
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"

ORIGIN
Query Match      3.2%; Score 79; DB 12; Length 867;
Best Local Similarity 45.7%; Pred. No. 4.4;
Matches 314; Conservative 0; Mismatches 370; Indels 3; Gaps 1;

QY 690 GACCGTGCCTGGAAGCTGAAGCCCGGATGAGCGGCCCAAGTGAAGCAAGTGGCCCT 749
DB 636 CGCCGCGTACATGAAGCTCTCAAGAGAGCAGGCGCGCCCTTCGACCGCCAGTTCTC 695
QY 750 GACCGAGAGAGATCAAGGCCCTGACCGCATCTGCGAGGAGATGGAAGAGGAGGCAA 809
DB 696 GCGCTACCGCAGAGGAGGTGTCACCGCGCATGCTCGAGAGATCTACACCGAGGCCA 755
QY 810 GATCACCAAGATCGGCCCGGAGAACCCCTACAACACCCCGCTTGGCATCAAGAGAA 869
DB 756 -----CAAGCAGATCCGCGCGGACCCACCTTGTCTCCCAAGCGCGCTCGAAGCCCGA 809
QY 870 GGACAGCAACCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAACAGCGCACCCAGGA 929
DB 810 GGGCGCCCAAGCCCAAGCACTGGGGCAAGCGGAGGTGAGTACCAAGAGCGCGAAGACCG 869
QY 930 CTTCTGGGAGTGCAGCTGGGCATCCCCACCCCGCGCTGAAGAGAGAGAGAGCGGT 989
DB 870 CGTGGCCCAAGAAAGGTCCGCTGCGCTACCGCGACGCCCGCCCAAGAGCGCAATATTC 929
QY 990 GACCGTGTGAGCTGGGCGAGCGCTACTTCAAGCTGCCCTGGACGAGG 1039
DB 930 CGGTGCTGTACACCGCATCTCCCGCTGCGCGCTTGGCTGCTNGCG 979

EST.
LOCUS      BM321430
DEFINITION Mastigamoeba balamuthi cDNA similar to ribosomal protein S4, mRNA
ACCESSION  BM321430
VERSION     BM321430.1 GI:18055836
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
ORGANISM    Mastigamoeba balamuthi
REFERENCE    1 (bases 1 to 867)
AUTHORS      Baptiste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
                Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
                Philippe,H.
TITLE        The analysis of 100 genes supports the grouping of three highly
                divergent amoebae: Dictyostellium, Entamoeba, and Mastigamoeba
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE      21819461
PUBMED       11830664
COMMENT      Contact: Muller Miklos
                Laboratory of Biochemical Parasitology
                The Rockefeller University
                1230 York Avenue, New York, NY 10021, USA
                Email: mmuller@rockvax.rockefeller.edu
                Insert Length: 867 Std Error: 0.00
                POLYA=No.

FEATURES             Location/Qualifiers
     source           1..867
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"

ORIGIN
Query Match      3.2%; Score 79; DB 12; Length 867;
Best Local Similarity 45.7%; Pred. No. 4.4;
Matches 314; Conservative 0; Mismatches 370; Indels 3; Gaps 1;

QY 58 CAGCGAGCAACTTCAAGGCCCGGAGCGCATCATCAAGTCTTCACTGCGCGAGGAG 117
DB 131 CCGCACAAGATGCGCGAGTGCCTGCGGTGATCATCTCTGTCGCGCAACAGGTTGAAGTAC 190
QY 118 GGCCACATCGCCCGCAACTGCCGCGCCCGCCCGCAAGAGGGTCTGTGAAGTSCGCGAAG 177

```



Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Saccharum.

1 (bases 1 to 671)

Vettore, A.L., da Silva, P.R., Kemper, E.L. and Arruda, P.

The libraries that made SUCEST

Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

Contact: Arruda P

Centro de Biologia Molecular e Engenharia Genetica

Universidade Estadual de Campinas

Caixa Postal 6010, 13083-970, Campinas SP, Brazil

Tel: 55 19 3788 1137

Fax: 55 19 3788 1089

Email: parruda@unicamp.br

Clone distribution: clone distribution information can be found

through the Brazilian Clone Collection Center (BCCC) at

<http://www.bcccenter.fcav.unesp.br>

Plate: 001 row: F column: 12

Seq primer: M13/Reverse primer.

Location/Qualifiers

1. .671

/organism="Saccharum officinarum"

/mol\_type="mRNA"

/db\_xref="taxon:4547"

/clone="SCCCL3001F12"

/lab\_host="XLBlue MRF"

/clone\_lib="CL3"

/note="Organ: Pool of sugarcane calli submitted to low

(40C) and high (37 C) temperature stress; Vector:

pBluscript; Site 1: EcoRI; Site 2: XhoI; An

unidirectional cDNA library generated from [pool of

sugarcane calli submitted to low (40C) and high (37 C)

temperature stress]. cDNA was prepared from polyA+ mRNA

using ZAP - cDNA Synthesis Kit (Stratagene). The

double-strand cDNAs were fractionated in a sepharose CL-2B

40cm-columns and fragments sizing between 0.8 and 1.5 Kb

were directionally cloned into the vector. Details of

each source of RNA and library construction can be

obtained at <http://sucet.lad.ic.unicamp.br/public>

FEATURES  
source

ORIGIN

Query Match 3.1%; Score 75.6; DB 13; Length 671;  
Best Local Similarity 47.3%; Pred. No. 11;  
Matches 228; Conservative 0; Mismatches 254; Indels 0; Gaps 0;

QY 986 GCGTGACCGTGTGGACGTGGGGGAGCCCTACTTTCAGGTGCGCCCTGGAGGAGACTTCC 1045  
DB 125 GCCTCAGCAGGTCTACGGCGCGCGGCGCATTCGGCGGCGCCCAATGCCGTGACTACG 184  
QY 1046 GCAAGTACACCGCTTCCACCATCCCGAGCATCAACAGAGACCCCGGCATCCGCTACC 1105  
DB 185 GCGGCGAGATGGGTTCTTGTAGTGTCTGACGCCAAGCGCGGGGGCGGCTCGATGAG 244  
QY 1106 AGTAAACGTGTGTCCTCCAGGGGTGAAGGGGAGCCCGAGCATCTTCCAGAGCAGCATGA 1165  
DB 245 GCGTCCCGCGGCGCTCGGGCTGTGACGTGTGAGGTGAGCGCGGCTAGTGGAGAGCGCGGTG 304  
QY 1166 CCAAGATCTGTGAGCCCTTCCGGCGCGCAACCCGAGATCGTATACAGTACATGG 1225  
DB 305 CGTTTGGCGCTGACCCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 364  
QY 1226 ACAGCTGTACGTGGGCGAGCAGCTGGAGATGGCGCAGACCCCGCGCGCGCGCGCGCG 1285  
DB 365 TCGAGGGGATGCGTCTCAAGAACCTGATGGCGGAGCGCGCGCGCGCGCGCGCGCGCG 424  
QY 1286 TCGGCGAAGACCTGTCTGGCTGGGCTTCAACACCCCGAGAGAGCAGCAGAGAGAGC 1345  
DB 425 ACCGCTCTCTCATGTCTGCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 484  
QY 1346 CCCCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1405  
DB 485 TCCTGGCGGACACGATCGACTACATGAAGAGCTGCTGGAGAGGATCAAGCTGCTGCTG 544  
QY 1406 AGCTGCCGAGAGAGAGAGCTGAGCCGTGAAAGAGATCCAGAGAGCTGCTGGGCAAGCTGA 1465

Db 545 AGGAGATCGACGACGACGAGGAGGCGCGCGCATGCTCAACGTTTCCGGAGCTCA 604  
QY 1466 AC 1467  
DB 605 AC 606

RESULT 9

AY106831

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. .2299

/organism="Zea mays"

/mol\_type="mRNA"

/db\_xref="taxon:4577"

/clone\_lib="Maize Mapping Project/DuPont Cornsensus Library"

/notes="this sequence is part of a project of EST

assemblies resulting from the application of public

contigs to seed DuPont contigs; this resource was

assembled by DuPont as part of a collaboration for the

overgo addressing of BACs in conjunction with the Maize

Mapping Project"

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

3.1%;

Score 75.4;

DB 11;

Length 2299;

Pred. No. 14;

0;

0;

936;

37;

6;

321

CGGGGCAACCCCGCAGCGAGCG

380

256

CGGCTCCGACGGGCGACACGCTCCCGGAGTTTCGCTGCGCGCGCGCGCGCGCGCGCGCGCG

315

381

GATCACCTCTGGCAGCG

440

316

CAACACCTCTCTCCAGGGCTACTCCGGCATCCGCTTCGAGATCCTCGAGGCCATCACCA

375

441

CTTCTGGAACCG

500

376

GCTCATCAACACCG

434

501

GAGCCCCAAGATGATCG

560

435

CGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG

494

QY 561 CTTGATCGAGATCTCGGCAAGAGGCGCATCGGCAACCGTGTGTGATCGGCCCAACCCCGGT 620  
Db 495 C---GTACGGTTGACGCGAGAGGTGGAGCGCGCGGAGCGTTTCAAGCGCGCGCAT 551  
QY 621 GAACATCATCGGCGCGACATGCTGACCCAGCTGGGCTGACCTGAACTTCCCCATCAG 680  
Db 552 CGAGGCGGCTTCTTCAAGCTCAACCCCAAGAGGCGCTGCGCATCGTCAACGCAAGTC 611  
QY 681 CCCCATCAGACCGTGCCTCGTGAAGCTGAAGCCCGGCAATGACGCGCCCAAGGTGAAGCA 740  
Db 612 CGTGGGTCGGGCTCGCGCATGCTGTGTTTCGACGCCAACGTCCTCGCGCTCTGTC 671  
QY 741 GTGGCCCTGACCGAGGAGAGATCAAGCGCTGACCGCCATCTCGGAGGATGGAGAA 800  
Db 672 CGAGTCTGTGTCGGGCTTCTGCGAGTGATGAACGGGAAGCCGAGTACAGGACCA 731  
QY 801 GGAGGCGAAGATCACCAAGATCGGCCCCGAGAACCCCTTACAACACCCCGCTGTTGCGCAT 860  
Db 732 CTTTACCACAGCTCAAGCACACCCCGGG-----TCCATCGAGGCGCGGC 779  
QY 861 CAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGACTTCCGCGAGCTGAACAGCG 920  
Db 780 CATCATGAGACATCTCTGATGGAGCTCTCTTATGAAGCAGCGCAAGGAGGTGAACGC 839  
QY 921 CACCCAGGACTTCTGGAGGTGACGTGGGATCCCCCAACCGCGCGCTGAAGAAGAA 980  
Db 840 CATGACCGCTGCTGAAGCGAAGCAGGACAGGTACGCGTCCGCACTGCGCGAGTG 899  
QY 981 GAAGAGGTGACGCTGTGGAGTGGGCGAGCGCTACTTACGCTGCGCCCTGGAGGAGA 1040  
Db 900 GCTCGGCCCCAGATCGAGGTCTATCCGCGCGCCCAAGTCCATCGAGCGAGAGTCAA 959  
QY 1041 CTTCCGCAAGTACACCGCTTCAACATCCCAAGCATCAACACGAGACCCCGGCGATCCG 1100  
Db 960 CTCGGTCAACGACACCCCTGTCATCGAGCTCCACGTCGCAAGCGCTGCACGGCGCAA 1019  
QY 1101 CTACAGATCAACGCTGTGCCCGAGGCTGGAAGGGAGCGCCAGCATCTTCCAGAGCAG 1160  
Db 1020 CTTCCAGGCGACCCCGATCGCGGTGTCATGTGACAAACCCCGCTCGCGCTCGCAACAT 1079  
QY 1161 CATGACCAAGATCTTGAGGACCTTCCGCGCGCGCAACCCGAGATCGTGATCTACCACTA 1220  
Db 1080 CGGAGCTCATGTCGCCAGTCTTCGAGTGTGCAAGGTTCTACAAACAGGGCT 1139  
QY 1221 CATGAGCAGCTGTAGTGGGAGCGA-----CTGAGATCGGCCAGCACCGGCCAA 1274  
Db 1140 GACCTCCAACTTGGCGGCGAGCGCAACCCGAGCCTGGACTACGGTTTCAAGGGCACCAG 1199  
QY 1275 GATCAGGAGCTGCGCAAGCAGCTGCTGGCTGGGCTTCAACCCCGCGCAAGAGCA 1334  
Db 1200 GATCGCCATGGCTCTACTGCTCGAGTCCAGTACTGGCCAAACCCCATCAACCA 1259  
QY 1335 -----CCAGAAGGAGCCCGCTTCTGTGGATGGGCTACGAGCTGCAACCCGCA 1385  
Db 1260 CTTCCAGAGCGGAGCAGCAACACGAGCAGTGAACCTCCCTCGGCTCGTCTCGCGAG 1319  
QY 1386 GTGAGCGTGCAGCCATCGAGCTGCCGAGAAGGAGCTGGACCGTGAACGACATCA 1445  
Db 1320 GAAGACCGCGAGGCGTGGACATCTCAAGTCTATGCTGCCATGCTACATGTTGCGCT 1379  
QY 1446 GAAGCTGTGGCAAGCTGAATGGGCGCAGCAGATCTACCCCGGATCAAGGTGCGCCA 1505  
Db 1380 GTGCAGGCGGTGGACCTCGCCACCTCGAGGAGAACCTCAAGAGCGCGCTCAAGAGCTG 1439  
QY 1506 GCTGTCAAGCTGCTGCGCGGCGCAAGCGCTGACCGCATCTGTCGCCCTGACCGAGGA 1565  
Db 1440 CTTGATGGCGGTGGCGAGAGGTGTGACCAACAGCTTCGCGCGGACCTCCACAGCGC 1499  
QY 1566 GCGGAGCTGTGAGCTGGCGAGAACCGCGAGATCTCTGGCGGAGCGCTGCGAGCGGTGTA 1625  
Db 1500 GCGCTTCAGCGAAGAGCCCTGCTCACCGCCATCGACCGGAGCGGTGTACGGGTACTA 1559

QY 1626 CTAGACCCCGCAGCAGGACCTGGTGGCCGAGATCCAGAGCAGGCGCACGACGAGTGGAC 1685  
Db 1560 CGACGACCCCTCGACGCGCAACTCGCCCTGTATGAAGAAGATTCGGGCGCTGTGGTGA 1619  
QY 1686 CTACAGATCTTACCAAGGAGCCCTTCAAGAACTCTGAAGACCGGCAAGTACGCAAGATCG 1745  
Db 1620 CCA-----CGCCCTCGCAGCGGCGAGGAGGAGCGCAGCGCTCGTGTCTC 1673  
QY 1746 CACCGCCACACACAGCAGTGAAGCAGTGAACGAGCGGCTGCGAGAGATCGCCATGA 1805  
Db 1674 CAAGATCAACAGTTTCGAGGAGGAGCTGCGGAGCGCTGCCACGAGATGGAGCGCG 1733  
QY 1806 GAGCATCTGTGATCTGGGCGAAGACCCCAAGTTCGCGCTGCCATCCAGAGAGGACCTG 1865  
Db 1734 CCGCTGGCTTCGAGACCGCGCGCGGATGCCAACAGATCAAGGAGAGCGGCTC 1793  
QY 1866 GAGACCTGTGGACCGACTACTGCGAGCGACCTGATCCCGAGTGGAGTTCGTGA 1925  
Db 1794 GTACCGCTGTACCGCTTTCATCCGCGAGACCTCGCGCGCTGTACCTGACCGCGAGAA 1853  
QY 1926 CACCCCGCCCTGGTGAAGCTGTGTACGAGTGGAGAGGAGCGCCATCATCGGCGCGA 1985  
Db 1854 GCTCAAGTCCCGCGGAGAGTGCACAAAGGTGTTCTGTGGCCCTCAGCGAGGCAAGCT 1913  
QY 1986 GACCTTCTAGTGGAGCGCGCGCCAA 2012  
Db 1914 CATCGACCCCATGCTCGACTGCTCAA 1940

## RESULT 10

BM321023 862 bp mRNA linear EST 03-JAN-2002  
LOCUS rockefeller.0.594 Mastigamoeba balamuthi lambda ZAP II Library  
DEFINITION Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (BC  
3.3.1.1), mRNA sequence.

ACCESSION BM321023.1 GI:18055429

VERSION EST.

KEYWORDS Mastigamoeba balamuthi

SOURCE Mastigamoeba balamuthi

ORGANISM Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.

REFERENCE 1 (bases 1 to 862)

AUTHORS Rapreste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W.,

Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and

Philippe, H.

TITLE The analysis of 100 genes supports the grouping of three highly

divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)

MEDLINE 21819461

PUBMED 11830664

COMMENT Contact: Muller Miklos

Laboratory of Biochemical Parasitology

The Rockefeller University

1230 York Avenue, New York, NY 10021, USA

Email: mmuller@rockefeller.edu

Insert Length: 862 Std Error: 0.00

POLYA=No.

FEATURES Location/Qualifiers

source 1..862

/organism="Mastigamoeba balamuthi"

/mol\_type="mRNA"

/strain="ATCC 30984"

/db\_xref="taxon:108607"

/clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"

/note="syn: Phreatamoeba balamuthi"

ORIGIN

Query Match 3.0%; Score 75; DB 12; Length 862;

Best Local Similarity 46.2%; Pred. No. 13;

Matches 283; Conservative 0; Mismatches 327; Indels 3; Gaps 1;

QY 1734 CGCAAGATCGCAGCCGCCACACGAGCTGAAGAGAGCTGACCGAGCGCTGCAGAA 1793



```

Db 220 GCCAAGATCCGCTGGTCTCTGTGCAACATCTTCTGACGAGACCAACGCGCGCGCG 279
QY 1794 GATCCCATGAGAGCATCTGATCTGGGGCAAGACCCCAAGTTCGGCTTCCCATCCCA 1853
Db 280 CATCGCGAGCGCGAGTCTCGGCTTCTGCTTGGAGGGCGAGAACCTCCAGGAGTACTG 339
QY 1854 GAAGAGACCTGGAGACCTGGTGGACCGATCTACGATCTGGCAGGCCACCTGGATCCCGAGTG 1913
Db 340 GGAGTGCACTTGAAGGCGCTGTGCTTCGGCCCTTACCGAGGACCTCAGANCAATCGTCA 399
QY 1914 GGAGTTCGTGAACACCCCGCTGGTGAAGCTGTGGTACCAGCTGGAGAGGAGGCCAT 1973
Db 400 CGAGCGGCTGACCGGACTCTAATGATCCCAAGGGGTTCCGGCCGAGGACACCCCAA 459
QY 1974 CATCGGCCCGAGACCTTCTTACGTGGACGGCGCGCCCAACCCGAGACCAAGATCGGCAA 2033
Db 460 GCTGCTGGAGACGACGAGGGGCTCGAGGAGGTGCGCTTCCCTCAACACGTGTCAAGCA 519
QY 2034 GCGCGGCTACGTACCGACCGGCGCGGCGGCGGAGAGATCGTAGCTGACCGAGACCACAA 2093
Db 520 GGTCCAGAGGAGACCGCGGCTTCTGGCAACAAGTCTTCCCGAGATCCCGGTGTAG 579
QY 2094 CAGAAGACCGAGTGTGACGAGGATCCAGTGGCCCTGAGACGACGCGGACGAGGTGAA 2153
Db 580 CGAGAGACGACGACTGCGGCTGTATGAGGCTGTACGAGCTGACCGCGACGCAAGCTGCT 639
QY 2154 CATCGTACCGACGACGAGTACGCGCTCGGCATC---ATCCAGCGCCGAGCCGACAAGAG 2210
Db 640 GTTCCCGCGCTCAACGTCACAGCTCTGTACCAAGAGCAAGTTTGAAACATCTACGG 699
QY 2211 CGAGAGCGAGTGTGTGAACGAGATCATCGAGCAGTGTATCAAGAGGAGAGGTGTACCT 2270
Db 700 TTCCGCGCCTACGCTCATCGAGGATCAAGCGCGGACCGACGCTGTGCTCGCGGCAA 759
QY 2271 GAGTGTGGTCCCGCCCAAGAGGATCGCGGCGACAGCAGCAGATCGACAGCTGTGTAG 2330
Db 760 GTCGCCGTGTGCGGGGCTACGGGACGTGGGCAAGGGCTGCGCGGAGTCTGCGCGG 819
QY 2331 CAAGGGCATCCGC 2343
Db 820 CCAGGGTCCGCG 832

```

```

RESULT 11
LOCUS BZ568300
DEFINITION pacs2-164_7463.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
ACCESSION BZ568300
VERSION BZ568300.1 GI:27201058
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

```

```

REFERENCE 1 (bases 1 to 914)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers
1..914
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"

```

## FEATURES

```

1..914
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"

```

## ORIGIN

```

Query Match 3.0%; Score 74.6; DB 28; Length 914;
Best Local Similarity 46.8%; Pred. No. 15;
Matches 270; Conservative 0; Mismatches 304; Indels 3; Gaps 1;

QY 652 CTGGGGTGCACCTGAACTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAG 711
Db 112 CTGGCAACACAGACGAGCGGCTGATCTTCAGCAGGCGGGCGCTTGAGGTGAGGCC 171
QY 712 CCGCGCATGAGCGGCGCCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAAGATCAAGGCC 771
Db 172 GGCAGCTTGGACAAACCGCCAGGCGACGCTCCAGGCGCCAGGGGACAAACAGCTGGGTATC 231
QY 772 CTGACCGCATCTGGAGGAGATGAGAGAGGAGGAGATCACCAGATCGGCCCGGAG 831
Db 232 GGTGGCGGCTGGACAAACAGGCGGCGCTGGAGACGCGGGCGCGGTAACTAGACTG 291
QY 832 AACCCCTAACAAACACCCCGCTGTTCGCCATCAAGAAAGAGGACAGACCAAGTGGCGCAAG 891
Db 292 CAGAGCGGCGAGCTCGACAAACGCGCGCGCGCTGCTCAACAGCGCCAAAGGTTGGCTG 351
QY 892 CTGGTGGACTTTCGGGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGAGCTGGCG 951
Db 352 AAGCTGCTCACCAGGCTGTTCGACAAACAGCGCGCGCTGACCCAGGCGAGTTCGCTGAA 411
QY 952 ATCCCGCCACCCCGCGCTGAAAGAAAGAGAGCTGACCGCTGTGACCGTGGCGCGAC 1011
Db 412 ATCCGCGCGGCGAGGCGCTGCGCAACAGCAGAGGCG---CATGTCTCGCGCTGGCGCG 468
QY 1012 GCTTACTTACGAGTGGCCCTGACGAGGACTTCCGCAAGTACACCGCTTACCATCCCC 1071
Db 469 GACACCGCATGCTCACCGCTGCACTTCGACACACAGGAGTGGCGGCTTACGCCAGCGCG 528
QY 1072 AGCATCAACAAACGAGACCCCGCATCCGCTACCAAGTACAAAGTGTGTCGCCAGGGCTGG 1131
Db 529 CTGCTCAGCTCGACGGCGAGCGCTTCTCAACAGGCGCGCGCGCGCGCGCGCGCG 588
QY 1132 AAGCGCAGCCCGAGCATTTCCAGAGCAGCATGACCAAGATCCTGAGAGCCCTTCGCGGCC 1191
Db 589 AAGTGGCGCGCGGCGCATCGACTTCAGCTGGCGCGCGCTGCGCAACCGCTTGGCC 648
QY 1192 CGCAACCCCGAGATCGTGTATCTACAGTACATGGAGC 1228
Db 649 AGTTGGAAGCGAGAGCGAGCTGCACCTGCGCGCGCG 685

```

## RESULT 12

```

LOCUS AY109500
DEFINITION Zea mays C1506_1 mRNA sequence.
ACCESSION AY109500
VERSION AY109500.1 GI:21213244
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 3134)
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whittitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
REFERENCE 2 (bases 1 to 3134)
AUTHORS Coe,E.H.

```

TITLE Direct Submission  
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA  
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, [maizemap.org](http://maizemap.org); ZmDB, [www.zmdb.iastate.edu](http://www.zmdb.iastate.edu); TIGR, [www.tigr.org](http://www.tigr.org); or NCBI, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov). When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: [www.zmdb.iastate.edu](http://www.zmdb.iastate.edu).

FEATURES  
source  
1. 3134  
/organism="Zea mays"  
/mol\_type="mRNA"  
/db\_xref="maizeDB:610364"  
/db\_xref="taxon:4577"  
/clone\_lib="Maize Mapping Project/DuPont Cornsensus Library"  
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed Dupont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

# ORIGIN

Query Match 3.0%; Score 74.4; DB 11; Length 3134;  
Best Local Similarity 40.6%; Pred. No. 20;  
Matches 963; Conservative 0; Mismatches 1347; Indels 61; Gaps 8;

QY 115 GAGGCCACATCGCCCGCACTGCGCGCCCGCCCGCAAGAGGGCTGCTGGAGTGGGC 174  
DB |||||  
QY 329 GAGAGCTGTACGGCTCGCACCTGCGAGCTGGAGGTGGAGCGCGGTCGAGACCGCGCG 388  
DB |||||  
QY 175 AAGGAGGGCCACCGATGAAGAGTGCACCGAGGGCCAGGCAACTTCTTCGCGAGGAC 234  
DB |||||  
QY 399 ACGAAGGACCGCGAGCGCTGGATCGAGCGAACCGTGCTGGTGGAGTGCAGGGGA 448  
QY 235 TTGGCCCTTCCCGAGGGCAAGCCCGGAGTTCCCGAGGAGCAACCGCGCCCAAGC 294  
DB |||||  
QY 449 AGCACCCGCTCAACTGCGAGCGCGCTGAGCGCGGCTG-ATGCAACACGGCTTCAACC 507  
QY 295 CCCACAGCGCGAGCTGAGGTGGCGGGGACAAACCCCGCAGCGAGCGCGGCGCGAG 354  
DB |||||  
QY 508 CGGGCGCGCTGCACTAGTGGCAACACCGCGCGCTGCGCGNNNNNATGCGCGAG 567  
QY 355 CGCCAGGAGCCCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCGCTGGTGAATC 414  
DB |||||  
QY 568 TGGACCGTGGAGGTGACGGGCTCTGAGGGCGCGCGCGGCTCACCATGGAGAGCTG 627  
QY 415 AAGTGGGGCGGCAGATCAAGGAGCCCTGTGGACACCGCGCGCGAGCACCGTGTG 474  
DB |||||  
QY 628 GCGCGGACTTCCCGCGCTGAGATCCCGTGGAGCTGGCTGGCGGGGCAACCGCGCG 687  
QY 475 GAGGAGATGAGCTCCCGCGCAAGTGGAGCCCAAGATGATCGGCGGATCGCGGGTTC 534  
DB |||||  
QY 688 AAGGAGCAGAAATGCTGAGCAGCGGTGGCTTCAACTGGGCGCGCGCGGCTGTCC 747  
QY 535 ATCAGGTGCGCAGTACGACAGATCTGTATCGAGATCGGCAAGAGAGCCATCGGC 594  
DB |||||  
QY 748 ACGNN 804  
QY 595 ACCGTGCTGATCGGCGCCACCGCTGAAACATCATCGGCGCGCAACATCTGACCCAGCTG 654  
DB |||||  
QY 805 CCGCGCAAGGGCGGCGCTCAACGTGTCTTCGAGGGCGCGAGNNNNNNNNNNNNNN 864  
QY 655 GCGTCAACCTGAACTTCCCATCAGCCCATCGAGACCGTGGCGCGGTGAAGCC 714  
DB |||||  
QY 865 NNNNNNNNNNTACGGCACCAAGCTCACGCGGAGTGGGCGCTGGACCCCGCGGAC 924  
QY 715 GGCATGGAGCGGCCCAAGGTGAAGCAGTGGGCCCTGACCGGAGGAGAGATCAAGGCCCTG 774  
DB |||||  
QY 925 ATCATGCTGCCTACATGAGAACCGCGAGCGCGTGTCTGCCGACCGAGCTTCCCGGTG 984

QY 775 ACCGCATCTCGAGGAGATGGAGAGAGGGCAAGATCATCCAGATCGGCCCGGAGAC 834  
DB |||||  
QY 985 CCGGTCAATATCCCGCGGTGCATCGGTGCCCGCATGGTCAAGTGGTCAAGCGC---ATC 1041  
QY 835 CCTTCAACACACCCCGTGTTCGCCATCAAGAAGAGAGACACCAAGTGGCGCAAGCTG 894  
DB |||||  
QY 1042 ATCGTCANNNNNCGAGTCCGACAAATTACTATTTCATAGGACAAACCGGTCTCGCG 1101  
QY 895 GTGGACTTCGCGAGCTGAAACAGGCGACCCAGGACTTCTGGGAGGTGCAGCTGGGATC 954  
DB |||||  
QY 1102 TCGCACTCGAGCGCGAGCTGCCAACGACAGAGCTGGTGGTCAAGCCGAGTACATC 1161  
QY 955 CCGCACCCCGCGCGCTCAAGAAAGAGAGCGGTGAACGCTGCTGGAGCTGGGCGACGCC 1014  
DB |||||  
QY 1162 ATCAAGAGCTGAACATAAATCTCGTGATAACGACCGCGGGCACGACGAGATCTGCGC 1221  
QY 1015 TACTTCAGGTGCCCTCGAGCGAGACTTCGCAAGTACACCGCTTACCATCCCCAGC 1074  
DB |||||  
QY 1222 ATCAAGAGCTTACACACAGCGCGCTTACACCATGAAGGATACGCTACTCCGGCGGC 1281  
QY 1075 ATCAACACAGAGACCCCGGATCGCTACCACTACACAGTACAGCTGCTGCCCGAGGTGAAG 1134  
DB |||||  
QY 1282 GSCAAGAGGTGACCGGGTGGAGTGCCTGAGCGCTGGACGCGCGGAGACATGCTGGTGTGC 1341  
QY 1135 GCGAGCCCGAGCATCTTCCAGAGCAGATGACCAAGATCTCTGGAGCCCTTCCGCGCCGC 1194  
DB |||||  
QY 1342 CACTTCGACCCCGGAGAGCCCAACAGTACGCAAGTACTGGTGTGGTCTTCTGG 1401  
QY 1195 AACCCCGAGATCGTCTATCCAGTACATGACGACCTGTACGTGGGCGAGCGACCTGGAG 1254  
DB |||||  
QY 1402 TCGTTCAGGTGGAGTCTCTGACCTGCTCGGCGCAAGAGATCGCGCTCCCGCATGG 1461  
QY 1255 ATCGGCCAGCAGCGCGCAAGATCGAGAGTGGCGCAAGACCTGCTGCTGGTGGGGTTC 1314  
DB |||||  
QY 1462 GACCAGTCTCAACACCCAGCGCGAGAGTCTATATGGAACCTCATGGGATGATGAAC 1521  
QY 1315 ACACCCCGCAAGAAACACCAAGAGGAGCCCGCTTCTGTGTGATGGGTAGAGCTG 1374  
DB |||||  
QY 1522 AACTGTGGTTCAAGGTGAAGTGAACGTGTGCTCGCAAGGGCGAGATCGGGTGG 1581  
QY 1375 CACCCCGCAAGTGGACCGTGCAGCCCATCGAC-----TGCCCGAGAGAG 1422  
DB |||||  
QY 1582 GTGTTTCGACACCCGAGCGAGCGCGCAACAGCCCGGGGTGGATGGCGCGCAGAAG 1641  
QY 1423 AGCTGACCGTGAAACACATCCAGAGCTGGTGGGAGAGCTGAATGGGCGCGACAGATC 1482  
DB |||||  
QY 1642 CACNN 1701  
QY 1483 TACCCCGCATCAAGTGGCGCGAGCTGTGCAAGCTGTGCGCGCGCAAGGCCCTGACC 1542  
DB |||||  
QY 1702 ATGAACACACCGAGCTCGGCAAGCAGTTTCAATGTCCGAGGTGCGCAAGCACGCTCG 1761  
QY 1543 GACATGTCCTTTCAGCGAGGAGCGAGCTGGAGCTGG-----CCGAGAACCGGAG 1596  
DB |||||  
QY 1762 CAGGAGTTCGGCTGGATCGCGTGCAGCGCCACGCTTACGACTGCACCAAGTTCCTCAAG 1821  
QY 1597 ATCTTCGCGAGCGCTGCAAGCGCTGTACTACGACCCAGCAAGAGACTGGTGGCCCGAG 1656  
DB |||||  
QY 1822 GACACCCCGCGCGCGCGACAGCATCTCTCATCAAGCGCGCACCGACTGACACGAGGAG 1881  
QY 1657 ATCGAAGACGAGGCGCACGACAGTGGACCTTACAGATCTACAGAGAGCTTACAGAGAGCC 1707  
DB |||||  
QY 1882 TTTCGAGCCATCCACTTCGACAAAGGCGCTCTCTGACACCTTACCGCATCGGCGAG 1941  
QY 1708 TTCAAGAACTGAAGACCGGCAAGTACGCAAGATGCGCACCGCCCGCACCAACACAGCTG 1767  
DB |||||  
QY 1942 CTCTATCCACCGGCGACCGGCTACAGTCCGACACTCGGTCCAGCGGGCTCCGCTCTG 2001  
QY 1768 AAGAGCTGACCGAGGCGCTGCAAGATTCGCCATGAGAGAGATCGTGTATCTGGGGCGAG 1827  
DB |||||  
QY 2002 TCGCACTTCGCGCCATCGCGAGGCGCTCAGGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2061

QY 1828 ACCCCCAAGTTCCGCTCCCATCCAGAGAG-----ACC 1863  
 Db 2062 AAGATCCACTGCCCTCCTGTCGGCAAGAGAGTGTCCCGAGCTCCGCTCTTCGGC 2121  
 QY 1864 TGGGAGACCTGTGTGACGACCTACTGCGAGCCACCTGATCCCGAGTGGAGTTCGTG 1923  
 Db 2122 TTCTCGCTGCGCTCCCGCAGCAGAGTGTGCGCTCCCATCCGCAAGCAGCATCTTCGT 2191  
 QY 1924 AACACCCCCGCCCTGGTGAAGTGTGTACAGCTGGAGAGAGGCCCATCATCGGGCC 1983  
 Db 2182 TCGCCAGCAATGAAGGAAGTGTGTATCGGGGCTACAGCCACGAGCATGTGTGAC 2241  
 QY 1984 GAGACTCTTACGTGCGAGCGGCCCAACCGGAGACCAAGATCGGCAAGCGCGGTAC 2043  
 Db 2242 GAGATCGGCCTTTCGACTCTCTCTCAAGGTCTACTTCAAGAGC---AGACCCCAAG 2298  
 QY 2044 GTGACCGACCGGGCGGCGAGAGATCGTAGCTGTGACCGAGACCCACCAAGAGAGACC 2103  
 Db 2299 TTCCCAACCGGGGCTCATGACCCAGTCTCGACTCGCTTCGCTCGGCGCTACATC 2358  
 QY 2104 GAGTGTGAGGCATCAGTGGCTTCAGAGCAGCGGCGAGAGGTGAACATGTGACC 2163  
 Db 2359 GAGTCAAGGGCGGCTTGGCCACGTGGAGTACACCGCGCGGAGCTTCGTTCATCAAC 2418  
 QY 2164 GACAGCCAGTACGCTCGGCTGATCATCCAGGCCCAAGAGCGAGCGAGCGAGCTG 2223  
 Db 2419 GCGAGCGCGCATCGAGCGGCTTCGCCATGATCTCGGCGGAAGTGGATCAAGGCC 2478  
 QY 2224 GTGACCGAGATCATCGAGCGCTGTCAAGAGGAGAGAGTGTACCTGAGTGGTGCC 2283  
 Db 2479 ATGTACCGAGATCATCGAGCGGCTTCGCGGACCAAGCGGAGGACCAACCGAGATGAC 2538  
 QY 2284 GCCCAAGGGCATCGGGGCAACAGCAGATCGAAGCTGGTGAGCAAGGGCATCGC 2343  
 Db 2539 CTCTGTACCGCAACCGAGGAGAGACATCTCTCCGCGAGAGCTGACCGGTG 2598  
 QY 2344 AAGTGTGTTCTCGAGCGCATCATCGCGGCTGATGATCTACAGTACATGAGAGAC 2403  
 Db 2599 GCGGCGGAGTACCCGAGAGGCTCAAGGTGTGTACGTATCGACCAAGTAAAGCGCCG 2658  
 QY 2404 CTGTACGTGGGCGGCGGCGCTAGGATCG 2434  
 Db 2659 CCGAGGGGTGAAGTACAGCTTGGGTTCG 2699

# RESULT 13 LOCUS

CK159613/c  
 FGAS041079 Triticum aestivum FGAS: TaLts5 Triticum aestivum cDNA,  
 mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE

CK159613  
 CK159613.1 GI:389985955  
 EST.  
 Triticum aestivum (bread wheat)

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.  
 1 (bases 1 to 889)

## REFERENCE AUTHORS

Allard, R.F., Crosby, W.B., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
 Genswein, B., Grat, R., Gulick, P., Hrycan, L.D., Laroche, A.,  
 Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,  
 Penniket, C., Roach, J.L. and Sarhan, F.

## TITLE JOURNAL COMMENT

Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
 Unpublished (2003)  
 Contract: Wm L Crosby  
 Bioinformatics  
 University of Saskatchewan, Department of Computer Science  
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
 Saskatchewan, S7N 5A9, Canada  
 Tel: 306 966 1769  
 Fax: 306 966 2033  
 Email: fgas\_estscs.usask.ca

This sequence is the direct result of the Base calling software  
 Phred (default parameters). It is the raw base calls. To aid in the  
 identification of the high quality insert the software Lucy  
 (default parameters) has been run on this sequence. Lucy identified  
 the region [123,668].  
 Plate: TaLts539 row: N column: 07.  
 Location/Qualifiers  
 1.889

## FEATURES source

/organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Wheat line PI 178383"  
 /db\_xref="taxon:4565"  
 /lab\_host="DH5 alpha"  
 /clone\_lib="Triticum aestivum FGAS: TaLts5"  
 /note="Organ: Crown; Vector: pGEM-T; SSH (suppression  
 subtractive hybridization) cDNA library from genotype  
 PI178383 cold hardened at 2 C for 21 days and 49 days  
 (equal amount of cDNA pooled together before subtraction,  
 tester) and subtracted against genotype Norstar cold  
 hardened at 2 C for 1 day (24 H) (driver). Modified Smart  
 cDNA (Clontech) priming and non-directional cloning"

## ORIGIN

Query Match 3.0%; Score 73.8; DB 14; Length 889;  
 Best Local Similarity 47.4%; Pred. No. 18;  
 Matches 251; Conservative 0; Mismatches 277; Indels 1; Gaps 1;

QY 753 CGAGGAGAGATCAAGGCCCTGACCCCATCTGCGAGGAGTGGAGAGGGGCAAGT 812  
 Db 674 CGACGACACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAG 615  
 QY 813 CACCAAGATCGGGCGGCGAGAACCCCTTACAAACCCCGCTGTCGCGCATCAAGAAAGGA 872  
 Db 614 CAGCACACAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACAA 555  
 QY 873 CAGCAACAGTGGCGGCAAGCTGGTGAATTCGCGAGCTGAAACAAGCGCACCGAGCTT 932  
 Db 554 CAACAACGACAAACAACAACGAGCAACCAACGACGACGACGACGACGACGACGACAA 495  
 QY 933 CTGGGAGGTGCGAGTGGGCTATCCCC-CACCCCGCGGCTGGAAGAGAGAGAGGTGA 991  
 Db 494 CAACAACACACCAACCAACCAACGAGCAACAACAACAACAACAACAACAACAACG 435  
 QY 992 CGGTGTGAGCGTGGCGGACGCTACTTTCAGCGTGCCTTGGACGAGGACTTCCGAGT 1051  
 Db 434 ACCGACGACAAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACAA 375  
 QY 1052 ACACCGCTTCACCTCCCGGATCAACAACGAGACCCCGGCTACCGTACCGATCA 1111  
 Db 374 ACAACGACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 315  
 QY 1112 ACGTGTGCGCCGAGGCTGGAGGGCGAGCCCGAGCATCTTCCAGAGCATGACCAAGA 1171  
 Db 314 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 255  
 QY 1172 TCCTGAGCGCTTCGCGCGCGCAACCCCGAGATCGTGTATCTTACAGTACATCGGACGC 1231  
 Db 254 AGAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACG 195  
 QY 1232 TGTACGTGGGCGGACCTTGGAGATCGGCGACGACCGCGGCCAAGATCGA 1280  
 Db 194 ACGAACACGATACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACAA 146

## RESULT 14 CC678788/c

LOCUS  
 DEFINITION  
 CC678788  
 genomic survey sequence.  
 CC678788  
 CC678788.1 GI:32083564  
 GSS.  
 Zea mays



QY	748	CTGACCGAGGAGAGATCAAGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGCC	807
Db	482	TCGCGTTACGCCAAGGAGGGTGTACCGCGCGACATGCTCGAGAAGATCTACACCGAGGCC	541
QY	808	AAGATCACCAGATCGGCCCGCCGAGAACCCCTACACACCCCGCTGTTGCCATCAAGAAG	867
Db	542	CA-----CAAGCAGATCCGCGCCGACCGACCTTCGTCCCGCAAGCGCGCTCGAAGCCC	595
QY	868	AAGGACAGCACCAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG	927
Db	596	GAGGGCGCCAGGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACACGAGCGCAAGAAC	655
QY	928	GACTTCTGGGAGGTGAGCTGGGCATCCCGCATCCCGCGCGCGCTGAAGAAGAGAGAGC	987
Db	656	CGCGTCGCCCCAGAGAAGGTCCGCTGGGCTACCCCGACGCGCCCGCAAGAGCCAGTAAATT	715
QY	988	GTGACCGTGTGTGAGCTGGCGGACGCGCTACTTCAGCG	1024
Db	716	CCCGTGCCCTGTACACNGCCATCTCCCGCTGGGGCG	752

Search completed: April 10, 2004, 16:16:19  
 Job time : 4157.08 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 02:53:16 ; Search time 6342.08 Seconds  
(without alignments)  
16873.640 Million cell updates/sec

Title: US-09-610-313-30  
Perfect score: 2469  
Sequence: 1 gtccagccaccatgccga.....gggctagccgggaattc 2469

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pla.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vit.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2469	100.0	2469	6	AX455914	AX455914 Sequence
2	2442.2	98.9	2463	6	AX455915	AX455915 Sequence
3	2415.4	97.8	2457	6	AX455916	AX455916 Sequence
4	2052	83.1	2312	6	BD263706	BD263706 Improved
5	2052	83.1	2312	6	AX373389	AX373389 Sequence
6	2051.2	83.1	9166	6	AX427930	AX427930 Sequence
7	2042.8	82.7	9788	6	AX427936	AX427936 Sequence
8	2028.8	82.2	9169	6	AX427931	AX427931 Sequence
9	2027	82.1	9194	6	AX427925	AX427925 Sequence
10	2027	82.1	12411	6	AX427927	AX427927 Sequence
11	2025.4	82.0	9194	6	AX427926	AX427926 Sequence
12	2025.2	82.0	2306	6	BD263704	BD263704 Improved
13	2025.2	82.0	2306	6	AX373387	AX373387 Sequence
14	2020.4	81.8	9785	6	AX427938	AX427938 Sequence
15	2018.6	81.8	9189	6	AX427921	AX427921 Sequence
16	2017	81.7	9167	6	AX427933	AX427933 Sequence
17	2017	81.7	9170	6	AX427928	AX427928 Sequence
18	2017	81.7	9782	6	AX427935	AX427935 Sequence
19	2017	81.7	9783	6	AX427934	AX427934 Sequence
20	2017	81.7	9792	6	AX427932	AX427932 Sequence
21	2015.6	81.6	4353	12	AF287353	AF287353 Synthetic
22	1999	81.0	3015	6	AX455946	AX455946 Sequence
23	1998.4	80.9	2300	6	BD263705	BD263705 Improved
24	1988.4	80.9	2300	6	AX373388	AX373388 Sequence
25	1982.4	80.3	3009	6	AX455987	AX455987 Sequence
26	1969.4	79.8	4319	6	BD263639	BD263639 Improved
27	1969.4	79.8	4319	6	AX373322	AX373322 Sequence
28	1958.8	79.3	4352	12	AF287352	AF287352 Synthetic
29	1907	77.2	2577	6	AX457088	AX457088 Sequence
30	1905.8	77.2	4343	6	AX306428	AX306428 Sequence
31	1901.8	77.0	4341	6	AX306429	AX306429 Sequence
32	1892.2	76.6	1977	6	AX455954	AX455954 Sequence
33	1879.2	76.1	7897	6	AX427923	AX427923 Sequence
34	1858	75.3	2305	6	BD263702	BD263702 Improved
35	1858	75.3	2305	6	AX373385	AX373385 Sequence
36	1837	74.4	1965	6	AX455952	AX455952 Sequence
37	1831.2	74.2	2299	6	BD263703	BD263703 Improved
38	1831.2	74.2	2299	6	AX373386	AX373386 Sequence
39	1825.6	73.9	3024	12	AF287355	AF287355 Synthetic
40	1807.6	73.2	9407	6	AX427937	AX427937 Sequence
41	1678.6	68.0	4307	6	AX302563	AX302563 Sequence
42	1678.6	68.0	4307	6	AX019132	AX019132 Sequence
43	1678.6	68.0	4307	6	BD130468	BD130468 Antiviral
44	1659.4	67.2	4307	6	BD268901	BD268901 Anti-vira
45	1659.4	67.2	4307	6	AX035453	AX035453 Sequence

ALIGNMENTS

RESULT 1  
AX455914  
LOCUS AX455914 2469 bp DNA linear PAT 06-JUL-2002  
DEFINITION Sequence 30 from Patent WO0204493.  
ACCESSION AX455914  
VERSION AX455914.1 GI:21714899  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.

REFERENCE 1  
AUTHORS zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,  
polypeptides and uses thereof  
JOURNAL Patent: WO 0204493-A 30 17-JAN-2002;





```

Db 1981 GCCGAGACCTTCTACGTGAGCGCGCCGCCCAACCGGAGACCAAGATCGGCAAGCGCGC 2040
QY 2041 TACGTGACCGACCGCGCGCGCGAGAGATCGTGAGCTGACCGAGACCAACCAAGAG 2100
Db 2041 TACGTGACCGACCGCGCGCGCGAGAGATCGTGAGCTGACCGAGACCAACCAAGAG 2100
QY 2101 ACCGAGTGCAGGCGCATCCAGCTGGCCCTCGAGGACAGCGGCGAGGTGAACATCGTG 2160
Db 2101 ACCGAGTGCAGGCGCATCCAGCTGGCCCTCGAGGACAGCGGCGAGGTGAACATCGTG 2160
QY 2161 ACCGAGGCGAGTACGCGCTGGGCGATCATCCAGGCGCGCGCGAGGCGAGCGAG 2220
Db 2161 ACCGAGGCGAGTACGCGCTGGGCGATCATCCAGGCGCGCGCGAGGCGAGCGAG 2220
QY 2221 CTGCTGAACAGATCATCGAGCGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGTG 2280
Db 2221 CTGCTGAACAGATCATCGAGCGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGTG 2280
QY 2281 CCCGCCCAAGAGGCGATCGCGGCAACGAGCAGATCGACAAGCTGTGTAGCAAGGCGATC 2340
Db 2281 CCCGCCCAAGAGGCGATCGCGGCAACGAGCAGATCGACAAGCTGTGTAGCAAGGCGATC 2340
QY 2341 CGCAAGGTGTCTTCCTGGACGCGATCGATGGCGGCGATCGTGTACCAAGTACATGGAC 2400
Db 2341 CGCAAGGTGTCTTCCTGGACGCGATCGATGGCGGCGATCGTGTACCAAGTACATGGAC 2400
QY 2401 GACCTGTACGTGGCGAGCGCGCGCTTAGGATCGATTAAAGCTTCCCGGCGTAGCAC 2460
Db 2401 GACCTGTACGTGGCGAGCGCGCGCTTAGGATCGATTAAAGCTTCCCGGCGTAGCAC 2460
QY 2461 GGTGAATTC 2469
Db 2461 GGTGAATTC 2469

RESULT 2
AX455915
LOCUS 2463 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 31 from Patent WO0204493.
ACCESSION AX455915
VERSION AX455915.1 GI:21714900
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE
1 zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.
AUTHORS Polynucleotides encoding antigenic hiv type c polypeptides,
TITLE polypeptides and uses thereof
JOURNAL Patent: WO 0204493-A 31 17-JAN-2002;
CHIRON CORPORATION (US); University of Stellenbosch (ZA)
FEATURES
Location/Qualifiers
source 1..2463
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975YM"

ORIGIN
Query Match 98.9%; Score 2442.2; DB 6; Length 2463;
Best Local Similarity 99.6%; Pred. No. 3e-247; 3; Indels 6; Gaps 1;
Matches 2460; Conservative 0; Mismatches 0;

QY 1 GTGACGCCACCATGGCCGAGGCGCATGAGCCAGGCCACCGAGCGCAACATCTGTATGCG 60
Db 1 GTGACGCCACCATGGCCGAGGCGCATGAGCCAGGCCACCGAGCGCAACATCTGTATGCG 60
QY 61 CGCAGCAACTTCAAGGCGCCCGAGGCGCATCATCAAGTGTCTCAACTGTGGGCGAGAGGCG 120
Db 61 CGCAGCAACTTCAAGGCGCCCGAGGCGCATCATCAAGTGTCTCAACTGTGGGCGAGAGGCG 120
QY 121 CACATGCCCGCAACTGCGCGCGCCCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180

```

```

Db 121 CACATGCCCGCAACTGCGCGCGCCCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180
QY 181 GGCACACAGATGAAGACTGACCGAGCGCCAGGCAACTTTCTTCGCGAGGACCTGCGCC 240
Db 181 GGCACACAGATGAAGACTGACCGAGCGCCAGGCAACTTTCTTCGCGAGGACCTGCGCC 240
QY 241 TTCCCGCAGGGGCAAGCGCGCGAGTTCCCGCAGGAGAGACCGCGCCCAAGCGCCAC 300
Db 241 TTCCCGCAGGGGCAAGCGCGCGAGTTCCCGCAGGAGAGACCGCGCCCAAGCGCCAC 300
QY 301 AGCGGCGAGTGCAGGTGCGCGCGAGCAACACCCCGCGAGGCGCGCGCGAGCGCGAG 360
Db 301 AGCGGCGAGTGCAGGTGCGCGCGAGCAACACCCCGCGAGGCGCGCGCGAGCGCGAG 360
QY 361 GGCACCGCTGAATTTCCCGCGAGATCACTCTGTGGCAGCGCCCGCTTGTGAGCATCAAG 420
Db 361 GGCACCGCTGAATTTCCCGCGAGATCACTCTGTGGCAGCGCCCGCTTGTGAGCATCAAG 420
QY 421 GSGCGCGAGATCAAGAGGCGCCCTGTGTGAGCAACCGCGCGCGAGCAACCGCTGTGAGGAG 480
Db 421 GSGCGCGAGATCAAGAGGCGCCCTGTGTGAGCAACCGCGCGCGAGCAACCGCTGTGAGGAG 480
QY 481 ATGAGCTGCGCGGCAAGTGGAGCGCAAGATGATCGCGGCGATCGCGGGCTTATCAAG 540
Db 481 ATGAGCTGCGCGGCAAGTGGAGCGCAAGATGATCGCGGCGATCGCGGGCTTATCAAG 540
QY 541 GTGCGCGAGTACGACCGAGATCTCGAGATCTCGGCAAGAGGCGCATCGGACCGCTG 600
Db 541 GTGCGCGAGTACGACCGAGATCTCGAGATCTCGGCAAGAGGCGCATCGGACCGCTG 600
QY 601 CTGATCGCGCGCGCGCGCGTGAACATCATCGCGCGCAACATGCTGACCGAGTGGCTGCG 660
Db 601 CTGATCGCGCGCGCGCGCGTGAACATCATCGCGCGCAACATGCTGACCGAGTGGCTGCG 660
QY 661 ACCCTGAATTTCCCGCATCAGCGCGCGCGTGAAGCGCGTGAAGCTGAAGCTGAAGCGCGCATG 720
Db 661 ACCCTGAATTTCCCGCATCAGCGCGCGCGTGAAGCGCGTGAAGCTGAAGCTGAAGCGCGCATG 720
QY 721 GACGCGCGCGCAAGTGAAGCGTGGCGCTGACCGAGGAGAGATCAAGCGCGCTGACCGCC 780
Db 721 GACGCGCGCGCAAGTGAAGCGTGGCGCTGACCGAGGAGAGATCAAGCGCGCTGACCGCC 780
QY 781 ATCTGCGAGGAGTGGAGAGGAGGCGCAAGATCAACCAAGATCGGCGCGCGAGAACCCCTAC 840
Db 781 ATCTGCGAGGAGTGGAGAGGAGGCGCAAGATCAACCAAGATCGGCGCGCGAGAACCCCTAC 840
QY 841 AACACCGCGTGTTCGCGCATCAAGAGAGGAGCAGCAGCAGTGGCGCAAGCTGGTGAC 900
Db 841 AACACCGCGTGTTCGCGCATCAAGAGAGGAGCAGCAGCAGTGGCGCAAGCTGGTGAC 900
QY 901 TTCCGCGAGTGAACAGCGCGCGCGCGTGAAGCGTGGGAGGTGAGCTGGGCGATCCCCAC 960
Db 901 TTCCGCGAGTGAACAGCGCGCGCGCGTGAAGCGTGGGAGGTGAGCTGGGCGATCCCCAC 960
QY 961 CCGCGCGCGTGAAGAGAGAGAGAGCGTGAAGCGTGGCGAGTGGCGCGAGCGCTACTTC 1020
Db 961 CCGCGCGCGTGAAGAGAGAGAGAGCGTGAAGCGTGGCGAGTGGCGCGAGCGCTACTTC 1020
QY 1021 AGCGTGGCGTGAAGAGAGAGAGAGCGTGGCGAGTGGCGAGTGGCGCGAGCGCTACTTC 1080
Db 1021 AGCGTGGCGTGAAGAGAGAGAGAGCGTGGCGAGTGGCGAGTGGCGCGAGCGCTACTTC 1080
QY 1081 AACGAGACCGCGCGCATCGCTACCAAGTGAAGTGGCTGGCGCGAGGCGTGAAGGCGAGC 1140
Db 1081 AACGAGACCGCGCGCATCGCTACCAAGTGAAGTGGCTGGCGCGAGGCGTGAAGGCGAGC 1140
QY 1141 CCGAGCATCTTCAGAGCAGATGAGCGAGTGGCGAGTGGCGCGAGTGGCGCGAGCGCGCG 1200
Db 1141 CCGAGCATCTTCAGAGCAGATGAGCGAGTGGCGAGTGGCGCGAGTGGCGCGAGCGCGCG 1200
QY 1201 GAGATCGTGAATCAACAGTACATGAGCAGCGCTGTGAGTGGCGAGCGAGCTGGGATCGGC 1260

```

Db 1201 GAGATCGTGATCTACCA-----GGCCCCCTGTACTGGGGCAGGACCTGGAGATCGGC 1254  
Qy 1261 CAGCACCGCGCAAGATCGAGGAGCTGCGAAGCACTGTGTGGCTGGGGCTTCACCAACC 1320  
Db 1255 CAGCACCGCGCAAGATCGAGGAGCTGCGAAGCACTGTGTGGCTGGGGCTTCACCAACC 1314  
Qy 1321 CCCGACAAGAGACACCAAGAGAGCCCTTCTGTGGATGGCTACAGAGCTGCACCC 1380  
Db 1315 CCCGACAAGAGACACCAAGAGAGCCCTTCTGTGGATGGCTACAGAGCTGCACCC 1374  
Qy 1381 GACAAGTGGACCGTGCAGCCCATCGAGTGCAGAGAGAGAGAGCTGGACCTGGAACGAC 1440  
Db 1375 GACAAGTGGACCGTGCAGCCCATCGAGTGCAGAGAGAGAGAGCTGGACCTGGAACGAC 1434  
Qy 1441 ATCCAGAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500  
Db 1435 ATCCAGAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494  
Qy 1501 CGCAGCTGTGCAAGCTGCTCGGGGGCCAAAGGCCCTGACCGACATCTGTGCCCTGACC 1560  
Db 1495 CGCAGCTGTGCAAGCTGCTCGGGGGCCAAAGGCCCTGACCGACATCTGTGCCCTGACC 1554  
Qy 1561 GAGGAGCGGAGCTGGAGCTGGCGAGAACCGGAGATCTTGGCGAGCCCGTGCACGGC 1620  
Db 1555 GAGGAGCGGAGCTGGAGCTGGCGAGAACCGGAGATCTTGGCGAGCCCGTGCACGGC 1614  
Qy 1621 GTGTACTACGACCCAGCAAGGACCTGTGTGCCAGATCCAGAGAGAGGCGACGACGAG 1680  
Db 1615 GTGTACTACGACCCAGCAAGGACCTGTGTGCCAGATCCAGAGAGAGGCGACGACGAG 1674  
Qy 1681 TGGACCTACGATCTACAGAGCCCTTCAAGACCTGAGACCTGGACCGGCAAGTACGCCAAG 1740  
Db 1675 TGGACCTACGATCTACAGAGCCCTTCAAGACCTGAGACCTGGACCGGCAAGTACGCCAAG 1734  
Qy 1741 ATGCGACCGGCCACCAACAGCGTGAAGCAGTGCACGAGGCGTGCAGAGATCGCC 1800  
Db 1735 ATGCGACCGGCCACCAACAGCGTGAAGCAGTGCACGAGGCGTGCAGAGATCGCC 1794  
Qy 1801 ATGGAGAGATCGTATCTGGGGCAAGACCCCAAGTTCGCGCTCCCATCCAGAGGAG 1860  
Db 1795 ATGGAGAGATCGTATCTGGGGCAAGACCCCAAGTTCGCGCTCCCATCCAGAGGAG 1854  
Qy 1861 ACCTGGGAGACCTGTGTGACCGACTACTTGGCAGGCGCACTGTGATCCCGAGTGGAGTTC 1920  
Db 1855 ACCTGGGAGACCTGTGTGACCGACTACTTGGCAGGCGCACTGTGATCCCGAGTGGAGTTC 1914  
Qy 1921 GTGAACACCCCTCTGTGTGAAGCTGTGTGACGCTGGAGAGAGGACCATCATCGGC 1980  
Db 1915 GTGAACACCCCTCTGTGTGAAGCTGTGTGACGCTGGAGAGAGGACCATCATCGGC 1974  
Qy 1981 GCGGAGACCTTCTAGTGGACGGCGCGCCAAACCGGAGACCAAGATCGGCAAGCGCGC 2040  
Db 1975 GCGGAGACCTTCTAGTGGACGGCGCGCGCCAAACCGGAGACCAAGATCGGCAAGCGCGC 2034  
Qy 2041 TACGTGACCGACCGGGCGCGGAGAGATGTGTGACCTGTGACGAGACCAACCAAGAG 2100  
Db 2035 TACGTGACCGACCGGGCGCGGAGAGATGTGTGACCTGTGACGAGACCAACCAAGAG 2094  
Qy 2101 ACCGAGCTGACGCGCATCAGCTGCGCTTGCAGGACAGCGGAGGAGGTGAACATCGTG 2160  
Db 2095 ACCGAGCTGACGCGCATCAGCTGCGCTTGCAGGACAGCGGAGGAGGTGAACATCGTG 2154  
Qy 2161 ACCGAGCAGCAGTACGCCCTGGGCAATCAATCAGGCGCGACCGCAAGAGCGAGAGCGAG 2220  
Db 2155 ACCGAGCAGCAGTACGCCCTGGGCAATCAATCAGGCGCGACCGCGCAAGAGCGAGAGCGAG 2214  
Qy 2221 CTGGTGAACAGATCATCAGCAGCTGTATCAAGAGAGAGAGGTGACTCTGAGCTGGGTG 2280  
Db 2215 CTGGTGAACAGATCATCAGCAGCTGTATCAAGAGAGAGAGGTGACTCTGAGCTGGGTG 2274  
Qy 2281 CCGCGCCCAAGGGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTGGAGCAAGGGCATC 2340  
Db 2275 CCGCGCCCAAGGGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTGGAGCAAGGGCATC 2334

Qy 2341 CGCAAGGTGCTGTTCTCTGGACGGCATCGATGGCGGCATCTGTATCTACCAAGTACATGGAC 2400  
Db 2335 CGCAAGGTGCTGTTCTCTGGACGGCATCGATGGCGGCATCTGTATCTACCAAGTACATGGAC 2394  
Qy 2401 GACCTGTACGTGGGACGGCGGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACC 2460  
Db 2395 GACCTGTACGTGGGACGGCGGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACC 2454  
Qy 2461 GGTGAATTC 2469  
Db 2455 GGTGAATTC 2463  
RESULT 3  
AX455916  
LOCUS 2457 bp DNA linear PAT 06-JUL-2002  
DEFINITION Sequence 32 from Patent WO0204493.  
ACCESSION AX455916  
VERSION AX455916.1 GI:21714901  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1  
AUTHORS zur Wegede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,  
JOURNAL polyptides and uses thereof  
Patent: WO 0204493-A 32 17-JAN-2002;  
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)  
FEATURES  
Location/Qualifiers  
1..2457  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="PR975YMMW"  
ORIGIN  
Query Match 97.8%; Score 2415.4; DB 6; Length 2457;  
Best Local Similarity 99.3%; Pred. No. 1.9e-244;  
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;  
Qy 1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACAGGCCCAACATCTGTATGAG 60  
Db 1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACAGGCCCAACATCTGTATGAG 60  
Qy 61 GCGAGCAACTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCACTGGCGCAAGGAGGC 120  
Db 61 GCGAGCAACTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCACTGGCGCAAGGAGGC 120  
Qy 121 CACATCGCCGCAACTGCGCGCCCCCGCCCAAGAGGGGTCTGGAAGTGGCGCAAGGAG 180  
Db 121 CACATCGCCGCAACTGCGCGCCCCCGCCCAAGAGGGGTCTGGAAGTGGCGCAAGGAG 180  
Qy 181 GGCACACAGATGAAGGACTGACCGAGCCAGGCCAACTTCTCCGAGAGACCTGGCC 240  
Db 181 GGCACACAGATGAAGGACTGACCGAGCCAGGCCAACTTCTCCGAGAGACCTGGCC 240  
Qy 241 TTCCCCCAGGGCAAGCCCGCGAGTTCCCGAGGAGAGAACCGCGCCCAACAGCCCCCACC 300  
Db 241 TTCCCCCAGGGCAAGCCCGCGAGTTCCCGAGGAGAGAACCGCGCCCAACAGCCCCCACC 300  
Qy 301 AGCGGAGCTGAGTGGCGCGCAACACCCCGCAGCGAGCGCGCGCCGAGGCGCCAG 360  
Db 301 AGCGGAGCTGAGTGGCGCGCAACACCCCGCAGCGAGCGCGCGCCGAGGCGCCAG 360  
Qy 361 GGCACCTGAACTTCCCCCAGATCACCTTGTGGACGCGCCCTGTGTGAGCATCAAGGTG 420  
Db 361 GGCACCTGAACTTCCCCCAGATCACCTTGTGGACGCGCCCTGTGTGAGCATCAAGGTG 420  
Qy 421 GCGCGCAGATCAAGAGGCGCTTGTGGACACCGCGCGCAGACACCGTGTGGAGGAG 480  
Db 421 GCGCGCAGATCAAGAGGCGCTTGTGGACACCGCGCGCAGACACCGTGTGTGGAGGAG 480

QY	481	ATGAGCTGCTCCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGGCGGTTTCATCAAG	540
DB	481		
QY	541	GTGGCCAGTAGACACAGATCTCTGATCGAGATCTCGGCAGAAAGCCCATCGGCACCGTG	600
DB	541		
QY	601	CTGATCGGCCCAACCCCGTGAAATCATCTCGGCGCAACATCTGATCAACAGCTGGCGTGC	660
DB	601		
QY	661	ACCTGAACTTCCCATCAGCCCCATCGAGACGTCCTCGTGAAGCTGAAGCCCGGCATG	720
DB	661		
QY	721	GACGGCCCCAAGGTGAAGCAAGTGGCCCTGACCGAGGAAGATCAAGGCCCTGACCGCC	780
DB	721		
QY	781	ATCTCGAGGAGATGGAGAAGGAGGCAAGATCACCAGATCGGCCCGCAGAACCCCTAC	840
DB	781		
QY	841	AACACCCCGTGTTCGCCATCAAGAAAGAGAGCAGCACTGCGGCCGAGAACCCCTAC	900
DB	841		
QY	901	TTCCGCGAGTGAACAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCAC	960
DB	901		
QY	961	CCGCGCGCTTGAAGAAAGAGAGCGTGACCGTGTGACCTGCGGCGAGCGCTACTTC	1020
DB	961		
QY	1021	AGCGTCCCTGGACGAGGACTTCGCAAGTACACCGCTTACCATTCCCGACATCAAC	1080
DB	1021		
QY	1081	AACGAGACCCCGGCATCCGTCAGTACAACGCTGTGCCCGCAGGCTGGAAAGGCGAG	1140
DB	1081		
QY	1141	CCGAGATCTTCCAGAGCAGCATCAGCAAGATCTGGAGCGCTTCCGCGCCCGCAACCCC	1200
DB	1141		
QY	1201	GAGATCGTATCTACAGTACATGGACGACTGTACCTGGGCGAGCACTGGAGATCGGC	1260
DB	1201		
QY	1261	CAGCACCGGCGCAAGATCGAGAGCTGGCGAAGCACTCTGCGCTGGGCTTCAACCAC	1320
DB	1261		
QY	1321	CCGCAAGAGACACCAAGAGGCGCCCTCTCTGTGGATGGGCTACGAGCTGCACCCC	1380
DB	1321		
QY	1381	GACAGTGGAACGTCAGCCCATCGAGTGCCTCGAAGAGAGAGCTGGAACGTGAACGAC	1440
DB	1381		
QY	1441	ATCCAGAGCTGGTGGGCAAGCTGAACCTGGGCGAGCCAGATCTACCCCGGCATCAAGGTG	1500
DB	1441		
QY	1501	CGCGAGCTGTGAAGCTGTGCGGGCGCCAGGCCCTGACCGACATCTGTCCTGACCC	1560
DB	1501		

RESULT 4  
BD263706  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS

SOURCE	synthetic construct	Db	541	CCGTGAAGCTGAACCGCGGATGACGCGCCCAAGGTCAAGCAAGTGGCCCTTACCGAGG	600
ORGANISM	synthetic construct				
REFERENCE	artificial sequences.	Qy	758	AGAAGATCAAGGCCCTGACGCCATCTCGAGGAGATGAGAGGAGGCAAGATCACCA	817
AUTHORS	1 (bases 1 to 2312)	Db	601	AGAAGATCAAGGCCCTGAGAGATCTGACCGAGATGAGAGGAGGCAAGATCAGCA	660
TITLE	Greer, C., Selby, M. and Walker, C.				
	Improved expression of HIV polypeptides and production of	Qy	818	AGATCGGCGCGAGAACCCCTTACCAACACCCCGCTGTTCGCGCATCAAGAAAGAGACAGCA	877
JOURNAL	virus-like particles	Db	661	AGATCGGCGCGAGAACCCCTTACCAACACCCCGCTGTTCGCGCATCAAGAAAGAGACAGCA	720
	Patent: JP 2002533124-A 73 08-OCT-2002;				
COMMENT	CHIRON CORP				
	OS Artificial Sequence	Qy	878	COAGTGGCGCAAGCTGGTGACTTTCGCGAGCTGAACRAAGCGCACCCAGGACTTCTGGG	937
	PN JP 2002533124-A/73	Db	721	CCAAGTGGCGCAAGCTGGTGACTTTCGCGAGCTGAACRAAGCGCACCCAGGACTTCTGGG	780
	PD 08-OCT-2002	Qy	938	AGGTGACGTGGGATCCCGGATCCCGGCGGCTTGAAGAAAGAGAGCGTGACCGTGC	997
	PF 30-DEC-1999 JP 2000591193	Db	781	AGGTGACGTGGGATCCCGGATCCCGGCGGCTTGAAGAAAGAGAGCGTGACCGTGC	840
	PR 31-DEC-1998 US 60/114495, 01-DEC-1999 US 60/168471 PI	Qy	998	TGGAAGTGGGCGACCGCTTACTTACAGCTGCGCGCTGAGAGGAGCTTCCGCAAGTACACCG	1057
	SUSAN BARNETT, JAN ZUR MEGEDE, INDRESH SRIVASTAVA, YING LIAN, PI	Db	841	TGGAAGTGGGCGACCGCTTACTTACAGCTGCGCGCTGAGAGGAGCTTCCGCAAGTACACCG	900
	KARIN HARTOG,	Qy	1058	CCTTCAACATCCCGAGCATCAACAGAGACCCCGGCGATCCCGTACAGTACCAAGTGC	1117
	PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER PC	Db	901	CCTTCAACATCCCGAGCATCAACAGAGACCCCGGCGATCCCGTACAGTACCAAGTGC	960
	C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC	Qy	1118	TGCGGAGGCTGGGAAGGCGAGCGCCAGCATCTTCCAGAGAGATGACCAAGATCTCTGG	1177
	C12N5/10,	Db	961	TGCGGAGGCTGGGAAGGCGAGCGCCAGCATCTTCCAGAGAGATGACCAAGATCTCTGG	1020
	PC C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC	Qy	1178	AGCCCTTCCGCGCGCAACCCCGGAGATCGTGATCTACAGTACATGACGACACCTGTACG	1237
	Description of Artificial Sequence: FS(-). protmod.RTopt(+) FH Key	Db	1021	AGCCCTTCCGCGCGCAACCCCGGAGATCGTGATCTACAGTACATGACGACACCTGTACG	1080
	Location/Qualifiers	Qy	1238	TGCGGAGGCTGGGAAGGCGAGCGCCAGCATCTTCCAGAGAGATGACCAAGATCTCTGG	1297
FT source	1. .2312	Db	1081	TGCGGAGGCTGGGAAGGCGAGCGCCAGCATCTTCCAGAGAGATGACCAAGATCTCTGG	1140
FT	Location/Qualifiers	Qy	1298	TGCGGAGGCTGGGAAGGCGAGCGCCAGCATCTTCCAGAGAGATGACCAAGATCTCTGG	1357
	1. .2312	Db	1141	TGCGGAGGCTGGGAAGGCGAGCGCCAGCATCTTCCAGAGAGATGACCAAGATCTCTGG	1200
	/organism="synthetic construct"	Qy	1358	GGATGGGCTACAGAGTGCACCCCGGAGAGTGCACCGCTGACCGCATTCAGAGTGCACCGGAG	1417
	/organism="genomic DNA"	Db	1201	GGATGGGCTACAGAGTGCACCCCGGAGAGTGCACCGCTGACCGCATTCAGAGTGCACCGGAG	1260
	/db_xref="taxon:32630"	Qy	1418	AGGAGAGTGGACCGTGAACGATCCAGAGCTGGTGGGCAAGCTGGAAGTGGGCGAGCC	1477
		Db	1261	AGGAGAGTGGACCGTGAACGATCCAGAGCTGGTGGGCAAGCTGGAAGTGGGCGAGCC	1320
		Qy	1478	AGATCTACCCCGGCGCATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCGCC	1537
		Db	1321	AGATCTACCCCGGCGCATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCGCC	1380
		Qy	1538	TGACCGAGATCGTGGCCCTTGAACGAGGCGCGAGCTGGAGTGGCGGAGACCGCGAGA	1597
		Db	1381	TGACCGAGATCGTGGCCCTTGAACGAGGCGCGAGCTGGAGTGGCGGAGACCGCGAGA	1440
		Qy	1598	TCCTGCGGAGCGCGTGCACCGCTGTACTACGACCCCGAGAGGAGGAGCTGTGGCGCGAGA	1657
		Db	1441	TCCTGCGGAGCGCGTGCACCGCTGTACTACGACCCCGAGAGGAGGAGCTGTGGCGCGAGA	1500
		Qy	1658	TCAGAGAGCGGCGCGAGCTGACCGAGTGAACCTTACAGATCTACAGAGGCGCTTCAAGAGCC	1717
		Db	1501	TCAGAGAGCGGCGCGAGCTGACCGAGTGAACCTTACAGATCTACAGAGGCGCTTCAAGAGCC	1560
		Qy	1718	TGAGAGCGCGAGTACCGCTGACCGAGTGCACCGCGCCACACCGAGCTGAGAGCGAGCTGA	1777
		Db	1561	TGAGAGCGCGAGTACCGCTGACCGAGTGCACCGCGCCACACCGAGCTGAGAGCGAGCTGA	1620
		Qy	1778	CCGAGGCGCGTGCAGAGATCGCCATCGAGAGCATTCGTGATCTTGGGCGAGAGCCCGCAAGT	1837
FEATURES					
source					
ORIGIN					
Query Match	83.1%; Score 2052; DB 6; Length 2312;				
Best Local Similarity	93.6%; Pred. No. 2.3e-206;				
Matches 2165; Conservative	0; Mismatches 135; Indels 12; Gaps 2;				
Qy	170 GCGGCAAGAGGCGCACAGATGAGGACTGCAACGAGCGCGAGGAGCTTCTTCGCG		229		
Db	1 GCGGCGGAGAGGACCAAAATGAAAGATTGCACTGAGAGAGGAGCTTAATTTTCGCG		60		
Qy	230 AGGACTTGGCTTCCCGGAGGCGAGGCGCGAGTTCCTCCAGGAGAGAGCGCGCA		289		
Db	61 AGGACTTGGCTTCTGCGAGGCGAGGCGCGAGTTCAGAGCGAGGAGCGCGCA		120		
Qy	290 ACAGCCCGAGCGCGAGCTGAGGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGG		343		
Db	121 ACAGCCCGAGCGCGAGCTGAGGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGG		180		
Qy	344 CCGGCGCGAGGCGCGAGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG		397		
Db	181 CCGGCGCGAGGCGCGAGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG		240		
Qy	398 GCGGCGCGAGGCGCGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG		457		
Db	241 GCGGCGCGAGGCGCGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG		300		
Qy	458 CCGAGACACCTGCTGGAGGAGATGAGCTGCGCGGCGGAGTGAAGCGCCAGAGATGATCG		517		
Db	301 CCGAGACACCTGCTGGAGGAGATGAGCTGCGCGGCGGAGTGAAGCGCCAGAGATGATCG		360		
Qy	518 GCGGCGCGGCGGCTTCAAGGTGGCGGAGTGAAGTGGCGGAGTGAAGTGGCGGAGTTCGCG		577		
Db	361 GCGGCGCGGCGGCTTCAAGGTGGCGGAGTGAAGTGGCGGAGTGAAGTGGCGGAGTTCGCG		420		
Qy	578 GCAAGAGGCGCATCGGCGCGGCTGATCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG		637		
Db	421 GCAAGAGGCGCATCGGCGCGGCTGATCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG		480		
Qy	638 ACATGCTGACCGAGCTGGGCTGACCTGAACTTCCCGATCAGCGCGGCGGCGGCGGCGG		697		
Db	481 ACCTGCTGACCGAGCTGGGCTGACCTGAACTTCCCGATCAGCGCGGCGGCGGCGGCGG		540		
Qy	698 CCGTGAAGCTGAAGCGCGGCGATGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG		757		

Db	1621	CCGAGCGCTGCAGAGGTGAGCACCAGAGCATCGTGTATCTGGGCAAGATCCCCAAGT	1680
Qy	1838	TCGCGCTCCCATCCAGAAAGAGACCTTGGGAGACCTGGTGAGCCGACTACTTGGCAGGCCA	1897
Db	1681	TCAAAGTCCCATCCAGAAAGAGACCTTGGGAGCCCTGGTGATGGAGTACTTGGCAGGCCA	1740
Qy	1898	CCTGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGGTACACAGC	1957
Db	1741	CCTGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGGTACACAGC	1800
Qy	1958	TGGAGAGGAGCCATCATTCGCGCCGAGACCTTCTTACGTGGACGGCGCCGCCAACCGCG	2017
Db	1801	TGGAGAGGAGCCATCGTGGCGCCGAGACCTTCTTACGTGGAGCGCGCCGCCAACCGCG	1860
Qy	2018	AGACCAAGATCGGCAAGCGCCGCTACGTGACCGACCGCGCGCGCGGAGATCGTAGGCC	2077
Db	1861	AGACCAAGTGGGCAAGCGCCGCTACGTGACCGACCGCGCGCGCGGAGATCGTAGGCA	1920
Qy	2078	TGACCGAGACCAACCAACAGAGACCGAGCTGACAGGCCATCCAGTGGCCCTGCAGGACA	2137
Db	1921	TCGCGGACACCAACCAACAGAGACCGAGCTGACAGGCCATCCAGTGGCCCTGCAGGACA	1980
Qy	2138	GGGCGAGGAGTGAACATCTGTACCGACAGCAGCAGTACGCCCTGGGCAATCATCCAGGCC	2197
Db	1981	GGGCGCTGGAGTGAACATCTGTACCGACAGCAGTACGCCCTGGGCAATCATCCAGGCC	2040
Qy	2198	AGCCCGACAAGAGCGAGCGAGCTGTGAACCAAGATCATCGAGCAGCTGATCAAGAAGG	2257
Db	2041	AGCCCGACAAGAGCGAGCGAGCTGTGAACCAAGATCATCGAGCAGCTGATCAAGAAGG	2100
Qy	2258	AGAAGGTGATCTAGCTGGTGGCCCGCCGACAGGGGATCGGGGCAACAGCAGATCG	2317
Db	2101	AGAAGGTGATCTAGCTGGTGGCCCGCCGACAGGGGATCGGGGCAACAGCAGATCG	2160
Qy	2318	ACAAGTGTGTAGCAAGGCAATCCCAAGGTGCTGTCTCGACGGCATCGATGGCGCA	2377
Db	2161	ACAAGTGTGTAGCAAGGCAATCCCAAGGTGCTGTCTCGACGGCATCGATGGCGCA	2220
Qy	2378	TCGTGATCTACAGTACATGAGCAGCTGTAGTGGGAGCGGGGCGCCCTAGGATCGATT	2437
Db	2221	TCGTGATCTACAGTACATGAGCAGCTGTAGTGGGAGCGGGGCGCCCTAGGATCGATT	2280
Qy	2438	AAAAGCTCCCGGGCTAGCACCGGTGAATC	2469
Db	2281	AAAAGCTCCCGGGCTAGCACCGGTGAATC	2312
LOCUS	AR373389	2312 bp	DNA
DEFINITION	Sequence 84 from patent US 6602705.	linear	PAT 18-DEC-2003
ACCESSION	AR373389		
VERSION	AR373389.1	GI:40075492	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 2312)		
AUTHORS	Barnett, S.W., Megede, J., Greer, C. and Selby, M.		
TITLE	Expression of HIV polypeptides and production of virus-like particles		
JOURNAL	Patent: US 6602705-A 84 05-AUG-2003;		
FEATURES	Location/Qualifiers		
source	1..2312		
	/organism="unknown"		
	/mol_type="genomic DNA"		
ORIGIN			
Query Match	83.1%; Score 2052; DB 6; Length 2312;		
Best Local Similarity	93.6%; Pred. No. 2.3e-206;		
Matches 2165; Conservative	0; Mismatches 135; Indels 12; Gaps 2;		
Qy	170	GGGCAAGGAGGGCCACCAGATGAAGAGTGTGACCGGCGGCGGCAACTTCTTCCGGC	229

Db	1	GGGCGCGAAGGACACCAATGAAGATTCACCTGAGAGACAGGCTAATTTCTTCCGCG	60
Qy	230	AGGACTGTGCTTCCCGCAGGGCAAGGCGCGGAGTTCCCGAGGAGCAAGACCGCGCA	289
Db	61	AGGACTGTGCTTCCCGCAGGGCAAGGCGCGGAGTTCCCGAGGAGCAAGACCGCGCA	120
Qy	290	ACAGCCCCACAGCGCGAGCTGACAGTGGCGGGG-----ACAAACCCCGCAGCGAGG	343
Db	121	ACAGCCCCACAGCGCGAGCTGACAGTGGCGGGGCGGAGCAACACGCTGAGCGAGG	180
Qy	344	CGGCGCGCAGCGCGCAGGGGACCCCTG-----AATTCCTCCCGAGATCAACCTGTGGCAGC	397
Db	181	CGGCGCGCAGCGCGCAGGGGACCCGTGAGCTTCAACTTCCCGAGATCAACCTGTGGCAGC	240
Qy	398	GGCCCTGTGTGAGCATCAAGTGGCGGCGCAGATCAAGAGGCGCCCTGTGGACACCGCGG	457
Db	241	GGCCCTGTGTGAGCATCAAGTGGCGGCGCAGCTCAAGAGGCGCTGTGGACACCGCGG	300
Qy	458	CGACGACACCCGTGTGGAGGATGAGCTTCCCGGCGAAGTGGAAAGCCCAAGATGATCG	517
Db	301	CGACGACACCCGTGTGGAGGATGAACTTCCCGGCGAAGTGGAAAGCCCAAGATGATCG	360
Qy	518	GGGCTATCGGCGCTTCAATCAAGTGGCGGCGCAGTACGACAGATCTGTGATCGAGATCTCGG	577
Db	361	GGGCTATCGGCGCTTCAATCAAGTGGCGGCGCAGTACGACAGATCTGTGATCGAGATCTCGG	420
Qy	578	GCAAGAGGCGCATCGGACACCGTCTGATCGGCGCCCAACCCCGTGAACATCATTCGCGCGCA	637
Db	421	GCCACAAGGCGCATCGGACACCGTCTGATCGGCGCCCAACCCCGTGAACATCATTCGCGCGCA	480
Qy	638	ACATGTGACCGAGCTGGGCTGACCCCTGAACTTCCCGATCAGCCCGCATCGAGACCGTGC	697
Db	481	ACCTGTGACCGAGATCGGCTGACCCCTGAACTTCCCGATCAGCCCGCATCGAGACCGTGC	540
Qy	698	CCGTGAAGCTGAAGCGCGCATCGGCGCCCAAGGTGAAGCAGTGGCCCTTGAACCGAGG	757
Db	541	CCGTGAAGCTGAAGCGCGCATCGGCGCCCAAGGTGAAGCAGTGGCCCTTGAACCGAGG	600
Qy	758	AGAAGATCAAGGCGCTGACCGCATCTCGAGAGATGAGAGAGGAGGAGAGATCAACA	817
Db	601	AGAAGATCAAGGCGCTGAGGATCTGACCGGAGATGAGAGAGGAGGAGAGATCAACA	660
Qy	818	AGATCGGCGCGGAGAACCCCTCAACAACCCCGTGTTCGCCATCAAGAAAGAGGACAGCA	877
Db	661	AGATCGGCGCGGAGAACCCCTCAACAACCCCGTGTTCGCCATCAAGAAAGAGGACAGCA	720
Qy	878	CCAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGG	937
Db	721	CCAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGG	780
Qy	938	AGGTGAGCTGGGATCCCGCCCGCGGCTGAAGAGAGAGAGAGAGCGTGACCGTGC	997
Db	781	AGGTGAGCTGGGATCCCGCCCGCGGCTGAAGAGAGAGAGAGAGAGCGTGACCGTGC	840
Qy	998	TGACGCTGGCGAGCGCTACTTTCAGCGTGCCTTCGACGAGGAGGACTTCCGCAAGTACACCG	1057
Db	841	TGACGCTGGCGAGCGCTACTTTCAGCGTGCCTTCGACGAGGAGGACTTCCGCAAGTACACCG	900
Qy	1058	CCCTTACCATCCCGAGCATCAACAAGAGACCCCGCGCATCCGCTACCAAGTACAAAGTGC	1117
Db	901	CCCTTACCATCCCGAGCATCAACAAGAGACCCCGCGCATCCGCTACCAAGTACAAAGTGC	960
Qy	1118	TGCCCGAGGCTGAGAGGAGCGCCAGCATCTTTCAGAGCAGCATGACCAAGATCTCTGG	1177
Db	961	TGCCCGAGGCTGAGAGGAGCGCCAGCATCTTTCAGAGCAGCATGACCAAGATCTCTGG	1020
Qy	1178	AGCCCTTCCCGCGCGCAACCCCGAGATCTGATCTTACAGTACATGAGGAGCTGTACG	1237
Db	1021	AGCCCTTCCCGAGGAGAGCCCGCATCTGATCTTACCAAGTACATGAGGAGCTGTACG	1080
Qy	1238	TGGCAGCGACTGTGAGATCGGCGCAGCACCGCGCGAAGATCGAGGAGCTCGGCAAGCAC	1297





QY 542 TCGCCAGTACGACAGATCTGTGATCTGAGATCTGTGGGCAAGAGCCATCGGACCGTGC 601  
 Db 3500 TGGGCCAGTACGACAGATCTGTGATCTGAGATCTGTGGGCAAGAGCCATCGGACCGTGC 3559  
 QY 602 TGATCGGCCCAACCCCGTGAACATCATCTGGCCGCAACATCTGACCCAGCTGGGCTGCA 661  
 Db 3560 TGGTGGGCCCAACCCCGTGAACATCATCTGGCCGCAACATCTGACCCAGATCGGCTGCA 3619  
 QY 662 CCTGAAATTCCCATCAGCCCATCAGCCCATCAGAGCCGCTGAGAGTGAAGCCCGGCTGG 721  
 Db 3620 CCTGAAATTCCCATCAGCCCATCAGAGCCGCTGAGAGTGAAGCCCGGCTGG 3679  
 QY 722 AGCGCCCAAGTGAAGTGAAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCA 781  
 Db 3680 AGCGCCCAAGTGAAGTGAAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGAGGAG 3739  
 QY 782 TCTGAGAGAGTGAAGAGAGAGGAGGAGATCAACCAAGATCGGCCCGGAGAACCCCTACA 841  
 Db 3740 TCTGACCGAGATGAAGAGAGAGGAGGAGATCAACCAAGATCGGCCCGGAGAACCCCTACA 3799  
 QY 842 ACACCCCGTGTTCGCTCATCAGAGAGAGGAGGAGATCAACCAAGTGGCGCAAGCTGTGACT 901  
 Db 3800 ACACCCCGTGTTCGCTCATCAGAGAGAGGAGGAGATCAACCAAGTGGCGCAAGCTGTGACT 3859  
 QY 902 TCCGCGAGTGAACCAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCGCCACC 961  
 Db 3860 TCCGCGAGTGAACCAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCGCCACC 3919  
 QY 962 CCGCGGCTGAAGAGAGAGAGGAGGAGTGAACCGTGTGAGAGTGGCGCAAGCTGTGACT 1021  
 Db 3920 CCGCGGCTGAAGAGAGAGAGGAGGAGTGAACCGTGTGAGAGTGGCGCAAGCTGTGACT 3979  
 QY 1022 GGTGCGCTTGACAGAGGACTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACA 1081  
 Db 3980 GGTGCGCTTGACAGAGGACTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACA 4039  
 QY 1082 ACAGACCCCGGCTACCGTACAGAGTCAAGTGTGCTGCGCCCGAGGCTGGAAGGCGAGCC 1141  
 Db 4040 ACAGACCCCGGCTACCGTACAGAGTCAAGTGTGCTGCGCCCGAGGCTGGAAGGCGAGCC 4099  
 QY 1142 CCAGATCTTCAGAGCAGCATGACCAAGATCTGAGGCGCTTCCGCGCGCAACCCCG 1201  
 Db 4100 CCAGATCTTCAGAGCAGCATGACCAAGATCTGAGGCGCTTCCGCGCGCAACCCCG 4159  
 QY 1202 AGATCGTGTATACAGTACATGAGCAGCTGTGAGTGGGAGCGACTGTGAGATCGGCC 1261  
 Db 4160 ACATCGTGTATACAGTACATGAGCAGCTGTGAGTGGGAGCGACTGTGAGATCGGCC 4219  
 QY 1262 AGACCCGCGCAGATCGAGGAGTGGCGAAGCAGCTGCTGGGCTTCAACACCC 1321  
 Db 4220 AGACCCGCGCAGATCGAGGAGTGGCGAAGCAGCTGCTGGGCTTCAACACCC 4279  
 QY 1322 CCACAAAGAGACCAAGAGGAGCCCGCTTCTGTGATGGCTACAGCTGCACCCCG 1381  
 Db 4280 CCACAAAGAGACCAAGAGGAGCCCGCTTCTGTGATGGCTACAGCTGCACCCCG 4339  
 QY 1382 ACAAGTGGACCGTGCAGCCCATCAGCTGCCGAGAGAGAGTGGACCGTGAACGACA 1441  
 Db 4340 ACAAGTGGACCGTGCAGCCCATCAGCTGCCGAGAGAGAGTGGACCGTGAACGACA 4399  
 QY 1442 TCAGAGCTGTGGGCAAGCTGAATGGGCGCAGCAGATCTACCCCGCAGTCAAGGTGC 1501  
 Db 4400 TCAGAGCTGTGGGCAAGCTGAATGGGCGCAGCAGATCTACCCCGCAGTCAAGGTGC 4459  
 QY 1502 GCCAGCTGTCAAGCTGTGCGGCGCGCAAGGCCCTGACCGACATCTGTCGCCCTGACCG 1561  
 Db 4460 GCCAGCTGTCAAGCTGTGCGGCGCGCAAGGCCCTGACCGACATCTGTCGCCCTGACCG 4519  
 QY 1562 AGAGGCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTGTGCGGAGCCCGTGCAGCGCG 1621  
 Db 4520 AGAGGCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTGTGAGGAGCCCGTGCAGCGCG 4579

QY 1622 TGTACTACGACCCCGACAGGACCTGTGTGGCCGAGATCCAGAGCAGGCGCCACCGACAGT 1681  
 Db 4580 TGTACTACGACCCCGACAGGACCTGTGTGGCCGAGATCCAGAGCAGGCGCCACCGACAGT 4639  
 QY 1682 GGAACCTACAGATCTTACGAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGA 1741  
 Db 4640 GGAACCTACAGATCTTACGAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGA 4699  
 QY 1742 TGGCACCACCCCGACACCAACGAGCTGAAGCAGCTGACCGAGCCCGTGCAGAAAGATCGCCA 1801  
 Db 4700 TGAAGGCGCCCGACACCAACGAGCTGAAGCAGCTGACCGAGCCCGTGCAGAAAGATCGCCA 4759  
 QY 1802 TGAGAGCATCTGTGATCTGGGCGAAGACCCCAAGTTCCGCTGCGCATCCAGAGAGGAGA 1861  
 Db 4760 CCAGAGCATCTGTGATCTGGGCGAAGACCCCAAGTTCAAGCTGCGCATCCAGAGAGGAGA 4819  
 QY 1862 CTTGGGAGACCTGTGTGAGCGGACTACTGTGCAGCGCCACCTGGATCCCGGAGTGGAGTTCCG 1921  
 Db 4820 CTTGGGAGGCTGTGTGAGCGGACTACTGTGCAGCGCCACCTGGATCCCGGAGTGGAGTTCCG 4879  
 QY 1922 TGAACACCCCGCTGTGTGTAACGCTGTGTAACGCTGTGTAACGCTGTGTAACGCTGTGTA 1981  
 Db 4880 TGAACACCCCGCTGTGTGTAACGCTGTGTAACGCTGTGTAACGCTGTGTAACGCTGTGTA 4939  
 QY 1982 CCAGAGCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGCGCT 2041  
 Db 4940 CCAGAGCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGCGCT 4999  
 QY 2042 ACCTGACCGACCCCGGCGCGGAGAGATCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAG 2101  
 Db 5000 ACCTGACCGACCCCGGCGCGGAGAGATCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAG 5059  
 QY 2102 CCAGCTGTGAGGCGCTTCCAGCTGGCCCTGCGAGAGACCGCGAGGCTGACATCGTGA 2161  
 Db 5060 CCAGCTGTGAGGCGCTTCCAGCTGGCCCTGCGAGAGACCGCGAGGCTGACATCGTGA 5119  
 QY 2162 CCAGAGCAGTACGCGCTTGGGCTATCATCCAGGCGCCAGCCGAGAGAGAGAGAGAGAGC 2221  
 Db 5120 CCAGAGCAGTACGCGCTTGGGCTATCATCCAGGCGCCAGCCGAGAGAGAGAGAGAGAGC 5179  
 QY 2222 TGGTGAACCAAGATCATCGAGCAGCTGATCAAGAGAGAGAGAGTGTACCTGAGTGGGTGC 2281  
 Db 5180 TGGTGAACCAAGATCATCGAGCAGCTGATCAAGAGAGAGAGAGTGTACCTGAGTGGGTGC 5239  
 QY 2282 CCGCCCAAGAGGCGCATCGGCGCAACGAGCAGATCCCAAGCTGTGTGAGCAAGGCGCATCC 2341  
 Db 5240 CCGCCCAAGAGGCGCATCGGCGCAACGAGCAGTGTGAGCGGCTGTGTGAGCGCGCATCC 5299  
 QY 2342 GCAAGTGTCTTCTTCTGAGCGGATCGA 2369  
 Db 5300 GCAAGTGTCTTCTTCTGAGCGGATCGA 5327

RESULT 7  
 LOCUS AX427936 9788 bp DNA linear PAT 20-JUN-2002  
 DEFINITION Sequence 174 from Patent WO0232943.  
 ACCESSION AX427936  
 VERSION AX427936.1 GI:21538023  
 KEYWORDS synthetic construct  
 SOURCE synthetic construct  
 ORGANISM synthetic construct  
 1 artificial sequences.  
 REFERENCE  
 AUTHORS Huang, Y. and Nabel, G.J.  
 TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for  
 Genetic immunization  
 JOURNAL Patent: WO 0232943-A 174 25-APR-2002;  
 GOVERNMENT OF THE UNITED STATES (US)  
 FEATURES  
 source Location/Qualifiers  
 1..9788  
 /organism="synthetic construct"  
 /mol\_type="unassigned DNA"



ORIGIN	/db_xref="taxon:32630"	Db	3915	CCGCCGGCTGAAGCAGAGAGAGCGTGCCTGGAGCTGGGCGACGCTACTTCA	3974
Query Match	82.7%; Score 2042.8; DB 6; Length 9788;	QY	1022	GGGTCCCTTGGACGAGACTTCCGCAAGTACACGGCTTCCACATCCCGAGCATCAACA	1081
Best Local Similarity	91.8%; Pred. No. 1.5e-205;	Db	3975	GGGTCCCTTGGACGAGACTTCCGCAAGTACACGGCTTCCACATCCCGAGCATCAACA	4034
Matches 2197; Conservative	0; Mismatches 177; Indels 18; Gaps 3;	QY	1082	ACGAGACCCCGGCTACCGCTACGAGTACAGTGTCTGCCCGAGGCTGGAGGCGACGC	1141
		Db	4035	ACGAGACCCCGGCTACCGCTACGAGTACAGTGTCTGCCCGAGGCTGGAGGCGACGC	4094
		QY	1142	CCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCTTCCGCGCCCGCAACCCG	1201
		Db	4095	CGCCCATCTTCCAGTGCAGCATGACCAAGATCTCTGGAGCTTCCGCGCAAGAGACCCG	4154
		QY	1202	AGATCGTGTATCTACAGATGATGAGCGAGCTGTAGTGGGAGCAGCTTGGAGATCGGC	1261
		Db	4155	ACATCGTGTATCTACAGATGATGAGCGAGCTGTAGTGGGAGCAGCTTGGAGATCGGC	4214
		QY	1262	AGCACCGCCCAAGATTCGAGGAGCTGCGCAAGCATCTGTGCGCTGGGCTTTCACCA	1321
		Db	4215	AGCACCGCCCAAGATTCGAGGAGCTGCGCGACACCTGTGCGCTGGGCTTTCACCA	4274
		QY	1322	CCGACAAAGAGCAACCAAGAGAGCGCCCTTCTGTGTGATGGCTACGAGCTGCAACCG	1381
		Db	4275	CCGACAAAGAGCAACCAAGAGAGCGCCCTTCTGTGTGATGGCTACGAGCTGCAACCG	4334
		QY	1382	ACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGGAACCGTGAAC	1441
		Db	4335	ACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGGAACCGTGAAC	4394
		QY	1442	TCAGAGAGCTGGTGGCAAGCTGAACTGGGCGAGCCAGATCTTACCCCGCATCAAGTGC	1501
		Db	4395	TCAGAGAGCTGGTGGCAAGCTGAACTGGGCGAGCCAGATCTTACCCCGCATCAAGTGC	4454
		QY	1502	GCCAGCTGTGCAAGCTGCTGCGCGCGCCAGAGCCCTGACCGACATCTGTCGCCCTT	1561
		Db	4455	GCCAGCTGTGCAAGCTGCTGCGCGCGCCAGAGCCCTGACCGACATCTGTCGCCCTT	4514
		QY	1562	AGAGGCGGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGCGAGCCGCTGACGG	1621
		Db	4515	AGAGGCGGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGCGAGCCGCTGACGG	4574
		QY	1622	TGTACTACGACCCCGAGAGGAGCTGGTGGCGCGAGATCCAGAGAGAGGCGGACGAGT	1681
		Db	4575	TGTACTACGACCCCGAGAGGAGCTGGTGGCGCGAGATCCAGAGAGAGGCGGACGAGT	4634
		QY	1682	GGACCTACAGATCTACAGAGAGCCCTTCAAGAACCTGAAACCGGCAAGTACGCGCGCA	1741
		Db	4635	GGACCTACAGATCTACAGAGAGCCCTTCAAGAACCTGAAACCGGCAAGTACGCGCGCA	4694
		QY	1742	TGCGCACCGCCACACCAACGAGCTGAGCAGCTGACCGAGGCGCTGACAGATCGCCA	1801
		Db	4695	TGAGGCGCGCCACACCAACGAGCTGAGCAGCTGACCGAGGCGCTGACAGATCGCCA	4754
		QY	1802	TGAGAGCATCTGATCTGGGCGCAAGACCCCGCAAGTTCCGCTTCCGCTCCAGAGAGA	1861
		Db	4755	CCGAGAGCATCTGATCTGGGCGCAAGACCCCGCAAGTTCCGCTTCCGCTCCAGAGAGA	4814
		QY	1862	CCTGGAGAGCATCTGATCTGGGCGCAAGACCCCGCAAGTTCCGCTTCCGCTTCCGCT	1921
		Db	4815	CCTGGAGAGCATCTGATCTGGGCGCAAGACCCCGCAAGTTCCGCTTCCGCTTCCGCT	4874
		QY	1922	TGAACACCCCGCTTGTGAGTGTGTGATACAGCTGGAGAGAGAGCCCATCATCGCG	1981
		Db	4875	TGAACACCCCGCTTGTGAGTGTGTGATACAGCTGGAGAGAGAGCCCATCATCGCG	4934
		QY	1982	CCGAGACCTTCTAGCTGAGCGCGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGCT	2041
		Db	4935	CCGAGACCTTCTAGCTGAGCGCGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGCT	4994
		QY	2042	AGTGTACCGAGCGCGCGCGCGAGAGATCTGTGAGCTTGACCGAGACCAACCAAGAGA	2101

Db	4995	ACGTGACCGACCGCGCGCCAGGAAGGTGGTGGCCCTGACCGACACACCAACCGAAGA	5051
Qy	2102	CCGAGCTGCAGGCGATCCAGCTGGCCCTGAGGACAGCGGCGAGGTTGAACATCGTGA	2161
Db	5055	CCGAGCTGCAGGCGATCCAGCTGGCCCTGAGGACAGCGGCGAGGTTGAACATCGTGA	5114
Qy	2162	CCGACAGCAGTACGCCCTGGGATCATCAGGCGCCAGCCCGACAGAGCGAGCGAGC	2221
Db	5115	CCGACAGCAGTACGCCCTGGGATCATCAGGCGCCAGCCCGACAGAGCGAGCGAGC	5174
Qy	2222	TGGTGAACCAAGATCATCAGCAGCTGATCAAGAAGAGAGAGGTGTACTCTCAGCTGGGTGC	2281
Db	5175	TGGTGAACCAAGATCATCAGCAGCTGATCAAGAAGAGAGAGGTGTACTCTCAGCTGGGTGC	5234
Qy	2282	CCGCGCCACAAGGGCATCGCGCGCCACAGCAGATCGACAAGCTGGTAGCAAGGGCATCC	2341
Db	5235	CCGCGCCACAAGGGCATCGCGCGCCACAGCAGATCGACAAGCTGGTAGCAAGGGCATCC	5294
Qy	2342	GCAAGGTGCTGTTCTCTCGACGGCATCGATGGCGCATCGTGATCTACCACTA	2393
Db	5295	GCAAGGTGCTGTTCTCTCGACGGCATCGATGGCGCATCGTGATCTACCACTA	5346
RESULT 8			
AX427931			
LOCUS	AX427931	9169 bp	DNA linear PAT 20-JUN-2002
DEFINITION	Sequence 169 from Patent WO0232943.		
ACCESSION	AX427931		
VERSION	AX427931.1	GI:21538018	
KEYWORDS	synthetic construct		
SOURCE	synthetic construct		
ORGANISM	artificial sequences.		
REFERENCE	1.		
AUTHORS	Huang, Y. and Nebel, G.J.		
TITLE	Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization		
JOURNAL	Patent: WO 0232943-A 169 25-APR-2002;		
	GOVERNMENT OF THE UNITED STATES (US)		
FEATURES	Location/Qualifiers		
source	1..9169		
	/organism="synthetic construct"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:32630"		
	/note="plasmid pVR1012x/s containing HIV genes"		
ORIGIN			
Query Match	82.2%;	Score 2028.8;	DB 6; Length 9169;
Best Local Similarity	91.7%;	Pred. No. 4.4e-204;	
Matches 2194;	Conservative 0;	Mismatches 182;	Indels 16; Gaps 4;
Qy	14	TGGCCGAGGCCATGAGCGCCACCGCGCCAAATCTCTGATGCGAGCGCAACTTCA	73
Db	2967	TGGCCGAGGCCATGAGCGGCCACCAATCATGATGCGCGGCAACTTCA	3026
Qy	74	AGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGGGCAAGAGGGCCACATCGCCGCA	133
Db	3027	AGGCCCCAAGCGCAT---CAAGTGTCTTCAACTGGGCAAGAGGGCCACTTGGCCGCA	3083
Qy	134	ACTCGCGCGCCCGCCGCAAGAAGGGCTGCTGGAAAGTGGCGCAAGAGGGCCACCAGATGA	193
Db	3084	ACTCGCGCGCCCTGGCGAAGAGGGCTGCTGGAAAGTGGCGCAAGAGGGCCACCAGATGA	3143
Qy	194	AGGACTGACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCTTTCGCCCAAGGCA	253
Db	3144	AGGACTGACCGAGCGCACAGGCTAA-TTTTTTAGGGAAGATCTGGCCCTTCCCACAAGGA	3202
Qy	254	AGGCCCGAGTTCCTCCAGCGAGGAGAACCGCGCCAAACAGCCCCACACCGCGCGAGTGC	313
Db	3203	AGGCCCGAGTATTTCTTCAGAGGAGACACAGAGCCCAAGCCCCACCAAGAGAGTCTTC	3262
Qy	314	AGGTGCGGG-----CGACAACCCCGCAGCGGCGCGGCGCGAGGCCAGGCA---364	





Db	3087	GCAACTGCGCGCCCGCCGCAAGAGGGCTGCTGGAAAGTGGCGCAAGGAGGGCCACCA	3146
Qy	191	TGAAGACTGCAACGAGCGCAGGCAACTTCTTCGCGAGACCTGGCTTCCCGCAGG	250
Db	3147	TGAAGGACTGCAACGAGCGCAGGCTAA-TTTTATTAGGGAAGATCTGGGCTTCCCA	3205
Qy	251	GCAAGGCCGCGAGTTCCCGAGGAGCAGAACCGCGCAACAGCCGCCACAGCGCGAGC	310
Db	3206	GGAAGGCCAGGGAATTTCTTCAGAGCAGACCAAGCCAAACAGCCGCCACAGAAAGAG	3265
Qy	311	TGAGAGTGGCGG-----CGACAAACCCCGCAGAGCGCGCGCGCGCGCGCGCGCA	364
Db	3266	TTGAGGTTTGGGGAAGAGACAACAACCTCCCTCTCAGAAAGCAGAGCGCATAGACA	3325
Qy	365	-----CCCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCGCTGGTGAGCATCA	418
Db	3326	CTGTATCTTTAGTCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCAACA	3385
Qy	419	TGGCGGCGAGATCAAGAGAGCCCTGTGGACACCGGCGCGGAGACACCGTCTGGAG	478
Db	3386	TAGGGGCGCAGCTGAAGAGGCGCTTTAGACACCGCGCGCGAGCACCGCTCTGGAG	3445
Qy	479	AGATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATCGGCGGCGATCGGCGGCTTCA	538
Db	3446	AGATGAACCTGCGCGGCGCTGGAAGCCCAAGATGATCGGCGGCGATCGGCGGCTT	3505
Qy	539	AGTGGCGCAGTACGACACAGATCTGATCGAGATCTGCGGCAAGAGCGCATCGGCA	598
Db	3506	AGTGGGCGCAGTACGACACAGATCTGATCGAGATCTGCGGCGCACAGGCGCATCG	3565
Qy	599	TGCTGATCGGCGCCACCGCTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGG	658
Db	3566	TGCTGTTGGGCGCCACCGCTGAAATCATCGGCGCGCAACCTGCTGACCCAGATCG	3625
Qy	659	GCACCTGAACTTCCCATCAGCCCCATCGAGACCGTGGCGGTGAAGTGAAGCGCGG	718
Db	3626	GCACCTGAACTTCCCATCAGCCCCATCGAGACCGTGGCGGTGAAGTGAAGCGCGG	3685
Qy	719	TGAGCGGCGCCAGGTGAGCAGTGGCGCTGAGCGCGCGCGAGGAGATCAAGCGCTG	778
Db	3686	TGAGCGGCGCCAGGTGAGCAGTGGCGCTGAGCGCGCGCGAGGAGATCAAGCGCTG	3745
Qy	779	CCATCTGAGGAGATGAGAGGAGGCGCAGATCAACAGATCGGCGCGCGAGAACCCCT	838
Db	3746	AGATCTGACCGAGATGAGAGGAGGCGCAGATCAACAGATCGGCGCGCGAGAACCCCT	3805
Qy	839	ACAACACCCCGTGTTCGCGATCAAGAGAGGAGCAGCAACAGTGGCGCAGCTGGTG	898
Db	3806	ACAACACCCCGTGTTCGCGATCAAGAGAGGAGCAGCAACAGTGGCGCAGCTGGTG	3865
Qy	899	ACTTCGCGAGTGAACAGCGCACCCAGGACTTCTGGAGGTGAGTGGGATCCCGCC	958
Db	3866	ACTTCGCGAGTGAACAGCGCACCCAGGACTTCTGGAGGTGAGTGGGATCCCGCC	3925
Qy	959	ACCCCGCGGCTGAGAGAGAGAGCGTGAACGCTGGTGGAGCGTGGGCGACGCTACT	1018
Db	3926	ACCCCGCGGCTGAGAGAGAGAGCGTGAACGCTGGTGGAGCGTGGGCGACGCTACT	3985
Qy	1019	TCAGGCTGCCCTGGAAGAGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCA	1078
Db	3986	TCAGGCTGCCCTGGAAGAGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCA	4045
Qy	1079	ACAACGAGACCCCGGCGATTCGCTACAGTACAGTGTGCTGCGCGAGGCTGGAGGCA	1138
Db	4046	ACAACGAGACCCCGGCGATTCGCTACAGTACAGTGTGCTGCGCGAGGCTGGAGGCA	4105
Qy	1139	GCCCGAGCATCTTCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCGCGCGCGCAAC	1198
Db	4106	GCCCGAGCATCTTCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCGCGCGCGCAAC	4165
Qy	1199	CGAGATCGTGATCTACAGTACATGAGACCTTGTAGCTGGGCGAGCGACCTGGAGATCG	1258

Db	4166	CCGACATCGTGAATCTACCAATACATGACGACCTGTGAGGAGCGACCTGGAGATCG	4225
Qy	1259	GCCAGCACCGCGCCCAAGATCGAGAGCTGGCGAAGCACCTGCTGCGTGGGGCTTCA	1318
Db	4226	GCCAGCACCGCACCAAGATCGAGAGCTGGCGAAGCACCTGCTGCGTGGGGCTTCA	4285
Qy	1319	CCCGCGACAGAGACCAACAGAGAGCCCGCTTCTGTGTGATGGGCTAGAGCTGCACC	1378
Db	4286	CCCGCGACAGAGACCAACAGAGAGCCCGCTTCTGTGTGATGGGCTAGAGCTGCACC	4345
Qy	1379	CCGACAACTGACCGTGTGAGCCCATCGAGCTGCGGAGAGAGAGAGCTGGACCGTGA	1438
Db	4346	CCGACAACTGACCGTGTGAGCCCATCGAGCTGCGGAGAGAGAGAGCTGGACCGTGA	4405
Qy	1439	ACATCCAGAACTGGTGGGCAAGCTGAATGGGCGACCGAGATACCCCGGCGATCAAG	1498
Db	4406	ACATCCAGAACTGGTGGGCAAGCTGAATGGGCGACCGAGATACCCCGGCGATCAAG	4465
Qy	1499	TGGCGCAGCTGTGAAGCTGTGCGCGCGCGCAAGGCGCTGACCGACATCGTGCCTGA	1558
Db	4466	TGGCGCAGCTGTGAAGCTGTGCGCGCGCGCAAGGCGCTGACCGACATCGTGCCTGA	4525
Qy	1559	CCGAGGAGCGGAGCTGGAGCTGGCGGAGAAACCGCGAGATCTCTGCGGAGCGCGTGC	1618
Db	4526	CCGAGGAGCGGAGCTGGAGCTGGCGGAGAAACCGCGAGATCTCTGCGGAGCGCGTGC	4585
Qy	1619	CGGTGATCTAGACCCCGCAGCAGGACCTGTGTGGCGGAGATCCAGAAAGCAGGCGCA	1678
Db	4586	CGGTGATCTAGACCCCGCAGCAGGACCTGTGTGGCGGAGATCCAGAAAGCAGGCGCA	4645
Qy	1679	AGTGGACCTACAGATCTACAGAGAGCGCTTCAAGAACTGAAAGACCGGCAAGTACG	1738
Db	4646	AGTGGACCTACAGATCTACAGAGAGCGCTTCAAGAACTGAAAGACCGGCAAGTACG	4705
Qy	1739	AGATCGCACCGCGCACACCGAGAGCTGAGAGAGCTGACCGAGCGCTGAGAGAGATCG	1798
Db	4706	GCATGAAGGCGCGCCACACCAACCGAGTGAAGAGCTGACCGAGCGCTGAGAGAGATCG	4765
Qy	1799	CCATGGAGAGCATCGTGATCTGGGCGAAGACCCCAAGTTCGCGCTGCCATCCAGAGG	1838
Db	4766	CCACCGAGAGCATCGTGATCTGGGCGAAGACCCCAAGTTCGCGCTGCCATCCAGAGG	4825
Qy	1859	AGACCTGGGAGACCTGTGTGAGACCGACTACTGGCGGCGCACCTGGATCCCGAGTGG	1918
Db	4826	AGACCTGGGAGACCTGTGTGAGACCGACTACTGGCGGCGCACCTGGATCCCGAGTGG	4885
Qy	1919	TCGTGAACACCCCGCGCGCTGTGTGAGAGCTGTGTGATCCAGTGGAGAGGAGGCGCAT	1978
Db	4886	TGTTGAACACCCCGCGCGCTGTGTGAGAGCTGTGTGATCCAGTGGAGAGGAGGCGCAT	4945
Qy	1979	GCGCGAGACCTTCTAAGTGAAGCGCGCGCGCAACCGCGAGACCAAGATCGGAGGCGG	2038
Db	4946	GCGCGAGACCTTCTAAGTGAAGCGCGCGCGCAACCGCGAGACCAAGATCGGAGGCGG	5005
Qy	2039	GCTAGTGAACCGCGCGCGCGCAAGAGATCGTGAGCTGACCGAGACCAACCAACAGAG	2098
Db	5006	GCTAGTGAACCGCGCGCGCGCAAGAGATCGTGAGCTGACCGAGACCAACCAACAGAG	5065
Qy	2099	AGACCGAGCTGAGCGCATTCAGTGTGGCGCTGAGGAGCAGCGGCGAGCGAGTGAACAT	2158
Db	5066	AGACCGAGCTGAGCGCATTCAGTGTGGCGCTGAGGAGCAGCGGCGAGCGAGTGAACAT	5125
Qy	2159	TGACCGAGCAGCGAGTACCGCTTGGCGCATCTCCAGGCGCGAGCGCGAGAGAGAGCG	2218
Db	5126	TGACCGAGCAGCGAGTACCGCTTGGCGCATCTCCAGGCGCGAGCGCGAGAGAGAGCG	5185
Qy	2219	AGCTGGTGAACCGAGTACCTGAGCAGCTGATCAAGAGAGAGAGTGTACTTGGCTGGG	2278
Db	5186	AGCTGGTGAACCGAGTACCTGAGCAGCTGATCAAGAGAGAGAGTGTACTTGGCTGGG	5245
Qy	2279	TGCGCGCGCGCAGGCGCATCGGCGCAGCAGCATCGACAAAGCTGTGTGAGAGAGGCA	2338
Db	5246	TGCGCGCGCGCAGGCGCATCGGCGCAGCAGCATCGGCGCAGGCTGTGTGAGAGAGGCA	5305

2339 TCCGCAAGGCTGTTCTCTGACGGCATCGATGCGGGCATCGTATCTACAGTA 2393  
 5306 TCCGCAAGGCTGTTCTCTGACGGCATCGAACAAGCCCGAGGAGCAGAGAA 5360  
 RESULT 11  
 AX427926  
 LOCUS AX427926 9194 bp DNA linear PAT 20-JUN-2002  
 DEFINITION Sequence 164 from Patent WO0232943.  
 ACCESSION AX427926  
 VERSION AX427926.1 GI:21538013  
 KEYWORDS  
 SOURCE synthetic construct  
 ORGNISM synthetic construct  
 artificial sequences.  
 1  
 REFERENCE  
 AUTHORS Huang, Y. and Nabel, G.J.  
 TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for  
 genetic immunization  
 JOURNAL Patent: WO 0232943-A 164 25-APR-2002;  
 GOVERNMENT OF THE UNITED STATES (US)  
 FEATURES  
 source Location/Qualifiers  
 i. 9194  
 /organism="synthetic construct"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32630"  
 /note="plasmid pVR1012x/s containing HIV genes"  
 ORIGIN  
 Query Match 82.0%; Score 2025.4; DB 6; Length 9194;  
 Best Local Similarity 91.6%; Pred. No. 1e-203;  
 Matches 2193; Conservative 0; Mismatches 186; Indels 16; Gaps 4;  
 14 TGGCCGAGGCCATGAGCAGG---CCACCAGGCGCAACATCTGATGAGCGCAGCAACT 70  
 2967 TGGCCGAGGCCATGAGCAGGTTGACCAAGCGGCCACCATCATGATGAGCGGCGCAACT 3026  
 71 TCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAATGCGGCGAAGAGGGCCACATCGCCC 130  
 3027 TCCGCAACCAAGCGCAAGATCGTGAAGTGTCTCAATGCGGCGAAGAGGGCCACACCGGCC 3086  
 131 GCAACTGCGCGCCCCCGCCAGAGAGGGCTGTGAGTGGCGGCAAGAGGGCCACCAAGA 190  
 3087 GCAACTGCGCGCCCCCGCCAGAGAGGGCTGTGAGTGGCGGCAAGAGGGCCACCAAGA 3146  
 191 TGAAGACTGACCGAGCGCCAGGCCAACTTCTCCGAGGACCTGGCTTCCGCCAGG 250  
 3147 TGAAGACTGACCGAGCGCACAGGCTAA-TTTTATAGGAAGATCTGGCTTCCACCAAG 3205  
 251 GCAAGGCCCGCGAGTTTCCCGAGCGAGCAAGACCGGCGCCAAAGCGCCCGCAGCGAGC 310  
 3206 GGAAGGCCAGGGAATTTCTTTCAGAGCAGACACAGAGCCCAAGCCCGCCAGAGAGAGC 3265  
 311 TGCAAGTGGCGCGG-----CGACAAACCCCGGAGAGCGCCCGGCGGAGCGCCAGGCA 364  
 3266 TTCAAGTTTGGGGAAGAGACAACACTTCCCTCTCAGAAAGCAGAGCGCGATAGACAAGAA 3325  
 365 -----CCCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCCCTGGTGGATCAAGG 418  
 3326 CTGTATCTTTCCTTTCCTTCAGATCACTCTTTGGCAGGAGCCCTCTGTCACAAATAAGA 3385  
 419 TGGGCGCGCAGATCAAGAGAGGCGCTGTGTGACACCGGCGCCGACGACACCGTGTGGAGG 478  
 3386 TAGGGGCGCAGCTGAAGAGGCGCCCTTCTAGACACCGGCGCCGACACACCGTGTGGAGG 3445  
 479 AGATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTCATCA 538  
 3446 AGATGAACCTGCGCGCGCGCTGGAAGCCCAAGATGATCGCGGCATCGCGGCTTCATCA 3505  
 539 AGGTGGCGCAGTACGACCAAGATCCTGTATCGAGATCTCGCGCAAGAGGCGCATCGCACCG 598  
 3506 AGGTGGCGCAGTACGACCAAGATCCTGTATCGAGATCTCGCGCAAGGCGCATCGGCACCG 3565

599 TGCTGATCGGCGCCACCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCT 658  
 3566 TGCTGTGGGCGCCACCCCGTGAACATCATCGGCGCGCAACCTGTGATGCCAGATCGGCT 3625  
 659 GCACCCCTGAATTTCCCATCAGCCCCCATCGAGACCGTGTGAGCTGAAGCCCGGCA 718  
 3626 GCACCCCTGAATTTCCCATCAGCCCCCATCGAGACCGTGTGAGCTGAAGCCCGGCA 3685  
 719 TGGACGGCGCCCAAGGTGAAGCAGTGGCCCTCAGCAGGAGAGATCAAGGCCCTGACCG 778  
 3686 TGGACGGCGCCCAAGGTGAAGCAGTGGCCCTCAGCAGGAGAGATCAAGGCCCTGAGG 3745  
 779 CCATCTGCGAGGAGATGGAAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCT 838  
 3746 AGATCTGCAACCGAGATGGAAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCT 3805  
 839 ACAACACCCCGCTGTCGCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 898  
 3806 ACAACACCCCGCTGTCGCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3865  
 899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGAGGTGACAGCTGGGCACTCC 958  
 3866 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGAGGTGACAGCTGGGCACTCC 3925  
 959 ACCCCCGCGCTGAAGAAG 1018  
 3926 ACCCCCGCGCTGAAGAAG 3985  
 1019 TCAGCGTGGCCCTGAG 1078  
 3986 TCAGCGTGGCCCTGAG 4045  
 1079 ACAACGAG 1138  
 4046 ACAACGAG 4105  
 1139 GCGCCGAGCATCTTCAG 1198  
 4106 GCGCCGAGCATCTTCAG 4165  
 1199 CCAGAGATCGTATCTACCAAGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1258  
 4166 CCAGCATCGTATCTACCAAGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4225  
 1259 GCAGAGACCGCGCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1318  
 4226 GCAGAGACCGCGCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4285  
 1319 CCGCCGAGCAAG 1378  
 4286 CCGCCGAGCAAG 4345  
 1379 CCAGAGATCGTATCTACCAAGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1438  
 4346 CCAGCATCGTATCTACCAAGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4405  
 1439 ACATCCAG 1498  
 4406 ACATCCAG 4465  
 1499 TGCGCCAG 1558  
 4466 TGCGCCAG 4525  
 1559 CCGAG 1618  
 4526 CCGAG 4585  
 1619 GCGTGTACTAG 1678  
 4586 GCGTGTACTAG 4645



1679	AGTGGACCTTACAGATCTTACAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCA	1738
QY		
4646	AGTGGACCTTACAGATCTTACAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCC	4705
DB		
1739	AGATGGCACCGCCCAACACAAAGAGTGAAGCAGCTGACCGAGGCGCTGCAGAAGATCG	1798
QY		
4706	GCATGAAGGGGCCCAACCAACGACGTGAAGCAGCTGACCGAGGCGCTGCAGAGATCG	4765
DB		
1799	CCATGGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCGCTGCCCATCCAGAAGG	1858
QY		
4766	CCACCGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCAAGCTGCCCATCCAGAAGG	4825
DB		
1859	AGACCTGGGAGACCTGCTGGACCGACTACTGGCAGGCGCACTGATGCCCGAGTGGGAGT	1918
QY		
4826	AGACCTGGGAGGCGCTGCTGGACCGACTACTGGCAGGCGCACTGATCCCCGAGTGGGAGT	4885
DB		
1919	TCGTGAACACCCCCCCTCGTGAAGCTGTGTACCAAGCTGGAGAAAGGACCCATCATCG	1978
QY		
4886	TCGTGAACACCCCCCCTCGTGAAGCTGTGTACCAAGCTGGAGAAAGGACCCATCATCG	4945
DB		
1979	GCGCCGAGACCTTCTACGTGGACGGCGCGCCCAACCGCGAGACCAAGATCGGCGAAGCGCG	2038
QY		
4946	GCGCCGAGACCTTCTACGTGGACGGCGCGCCCAACCGCGAGACCAAGCTGGCGCAAGCGCG	5005
DB		
2039	GCTACGTGACCGACCGGGGCGGCAGAGAAGATCGTGAGCCTGACCGAGACCAACCAACGAGA	2098
QY		
5006	GCTACGTGACCGACCGCGCGCGCCAGAAAGGTGGTGCCCTTGACCGACCAACCAACGAGA	5065
DB		
2099	AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAAGACAGCGGCAGCGAGGTGAACATCG	2158
QY		
5066	AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAAGACAGCGGCCTGGAGGTGAACATCG	5125
DB		
2159	TGACCGACAGCCAGTAGCCCTGGGCATCATCCAGGCGCCAGCCCGACAAGACGAGAGCG	2218
QY		
5126	TGACCGACAGCCAGTAGCCCTGGGCATCATCCAGGCGCCAGCCCGACAAGACGAGAGCG	5185
DB		
2219	AGCTGGTGAACAGATCATCAGCAGCTGATCAAGAAAGGAGAGGTGTACTGAGCTGGG	2278
QY		
5186	AGCTGGTGAACAGATCATCAGCAGCTGATCAAGAAAGGAGAGGTGTACTGAGCTGGG	5245
DB		
2279	TGCCCGCCCAACAAGGGCATCGCGCAACGACGAGATCGACAAGCTGGTGAGCAAGGCGCA	2338
QY		
5246	TGCCCGCCCAACAAGGGCATCGCGCAACGACGAGGTGACGCGCTGGTGAGCGCCGCGCA	5305
DB		
2339	TCGCCAAGGTGCTGTTCTGTGAACGGCATTCGATGGCGGCATCGTGATCTTACAGTGA	2393
QY		
5306	TCCGCCAAGGTGCTGTTCTGTGAACGGCATTCGACAAGGCGCCAGGAGGAGCAAGAGAA	5360
DB		

RESULT 12	
BD263704	
LOCUS	2306 bp DNA linear PAT 17-JUL-2003
DEFINITION	Improved expression of HIV polypeptides and production of virus-like particles.
ACCESSION	BD263704
VERSION	BD263704.1 GI:33073472
KEYWORDS	JP 2002533124-A/71.
SOURCE	synthetic construct
ORGANISM	artificial sequences.
REFERENCE	1 (bases 1 to 2306)
AUTHORS	Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.
TITLE	Improved expression of HIV polypeptides and production of virus-like particles
JOURNAL	Patent: JP 2002533124-A 71 08-OCT-2002; CHIRON CORP
COMMENT	OS Artificial Sequence PN JP 2002533124-A/71 PD 08-OCT-2002 PF 30-DEC-1999 JP 2000591193 PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 PI SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI

KARIN HARTOG,									
PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER, PC									
C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC									
C12N5/10,									
PC C12N7/02, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC									
Description of Artificial Sequence: fs(-).protmod.RTopt.YM FH Key									
Location/Qualifiers									
FT	source	1..2306	/organism=	'Artificial Sequence'.					
FT	Location/Qualifiers								
1..2306									
/organism="synthetic construct"									
/mol_type="genomic DNA"									
/db_xref="taxon:32630"									
ORIGIN									
Query Match		82.0%; Score 2025.2; DB 6; Length 2306;							
Best Local Similarity		93.3%; Pred. No. 1.5e-203;							
Matches 2156; Conservative 0;		Mismatches 138; Indels 18; Gaps 3;							
Qy	170	GGGCAAGGAGGGCCACGAGTGAAGACTGACCGAGCGCCAGAGCGCAACTCTTCCGCG	229						
Db	1	GGGCGCGGAAGACACCAATGAAGATTGCTAGAGACGGGTAAATTCTTCCGCG	60						
Qy	230	AGGAACTTGGCTTCCCCAGGCGCAAGCCCGGAGTTCGCCAGCAGCAGAACCCGCGCA	289						
Db	61	AGGAACTTGGCTTCTTCGAGGCGCAAGCCCGGAGTTCAGCAGCAGCAGACCCGCGCA	120						
Qy	290	ACAGCCCCACAGCGCGAGCTGCAGTGGCGGCGG-----ACAAACCCCGCAGCGAGG	343						
Db	121	ACAGCCCCACCCGCGGAGCTGCAGTGTGGGCGCGGAGAACAAACAGCTTGAGCGAGG	180						
Qy	344	CGGCGCGCAGCGCGCAGGCGACCCCTG-----AACTTCCCCAGATCACCTGTGGCAGC	397						
Db	181	CGGCGCGCAGCGCGCAGGCGACCGTGAGCTTCAACTTCCCCAGATCACCTGTGGCAGC	240						
Qy	398	GCCCCCTGGTAGCATCAAGTGGCGGCGCAGATCAAGAGGCGCTGTGGACACCGGCG	457						
Db	241	GCCCCCTGGTAGCATCAGGATCGCGGCGCAGCTCAAGAGGCGGTGTCTCGACACCGGCG	300						
Qy	458	CGCAGCACACCGTGTGTGAGGAGATGAGCTGCGCGGCAAGTGGAAAGCCCAAGATGATCG	517						
Db	301	CGCAGCACACCGTGTGTGAGGAGATGAACCTGCTCCGCGCAAGTGGAAAGCCCAAGATGATCG	360						
Qy	518	GCGCATCGCGGCTTTCATCAAGTGGCGCGCAGTACGACGAGATCTCTGATCGAGATCTGCG	577						
Db	361	GCGGATCGGCGGCTTTCATCAAGTGGCGCGCAGTACGACGAGATCTCCCGTGGAGATCTCGC	420						
Qy	578	GCAAGAGGCCATCGGCACCGTGTGATCGGCGCCACCCCGGTGAACATCATCGCGCGCA	637						
Db	421	GCCAAAGGCCATCGGCACCGTGTGTGTGGGCGCCACCCCGGTGAACATCATCGCGCGCA	480						
Qy	638	ACATGTGTACCCAGCTGGGCTGCAACCTGTAACTTCCCCATCGACCCCATCGAGACCGGTGC	697						
Db	481	ACCTGTGAGCCAGATGGGCTGCACCTGTAACTTCCCCATCGACCCCATCGAGACCGGTGC	540						
Qy	698	CGGTGAGCTGAGCCCGGCGATGACCGGCCCGCAGGTGAAGCAGTGGCCCTGACCGAGG	757						
Db	541	CGGTGAGCTGAGCCCGGCGATGACCGGCCCGCAGGTGAAGCAGTGGCCCTGACCGAGG	600						
Qy	758	AGAAGATCAAGGCCCTGACGCCATCTGCGAGGAGATGGAGAGGAGGCGCAAGATCACCA	817						
Db	601	AGNAGATCAAGGCCCTGTGTGAGATCTGCACCGAGATGGAGAGGAGGCGCAAGATCAGCA	660						
Qy	818	AGATCGGCCCGAGAACCCCTTAAACAACCCCGTGTGTGGCATCAAGAGAGAGGACAGCA	877						
Db	661	AGATCGGCCCGAGAACCCCTTAAACAACCCCGTGTGTGGCATCAAGAGAGAGGACAGCA	720						
Qy	878	CCAAGTGGCGCAAGCTGTGTGACTTTCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGG	937						
Db	721	CCMAAGTGGCGCAAGCTGTGTGACTTTCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGG	780						
Qy	938	AGGTGCAGCTGGGCACTCCCCACCCCGCGGCTGTGAAGAGAGAGAGCGCTGACCGGTGC	997						





Db	241	GC	CC	CC	CT	TG	T	G	A	C	A	C	A	T	C	A	G	A	T	C	G	C	G	C	C	A	G	T	C	A	A	G	A	G	C	G	C	T	G	C	A	C	A	C	G	G	C	300		
Qy	458	CC	G	A	C	G	A	C	A	C	G	T	C	T	G	A	G	A	G	A	T	A	G	C	T	C	C	G	C	G	A	A	G	T	G	A	A	C	C	A	A	G	A	T	A	T	C	G	517	
Db	301	CC	G	A	C	G	A	C	A	C	G	T	C	T	G	A	G	A	G	A	T	A	A	C	T	T	C	C	G	C	G	A	A	G	T	G	A	A	C	C	A	A	G	A	T	A	T	C	G	360
Qy	518	GC	G	C	A	T	C	G	G	G	T	T	C	A	T	A	A	G	T	G	C	C	A	T	A	G	C	A	C	A	G	A	T	C	C	T	G	A	T	C	G	A	T	C	G	577				
Db	361	GC	G	G	A	T	C	G	G	G	T	T	C	A	T	A	A	G	T	G	G	G	A	T	A	C	G	A	T	A	C	G	A	T	A	C	G	A	T	C	C	C	G	A	T	C	G	420		
Qy	578	GC	A	A	A	G	G	C	C	A	T	C	G	A	C	C	G	T	G	A	T	C	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	637				
Db	421	GC	C	A	A	G	G	C	C	A	T	C	G	A	C	C	G	T	G	T	G	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	480						
Qy	638	A	C	A	T	G	C	T	G	A	C	C	A	G	T	G	B	G	T	G	C	A	C	C	T	G	A	A	C	T	T	C	C	C	A	T	C	A	G	C	C	C	C	A	T	C	G	697		
Db	481	A	C	C	T	G	C	T	G	A	C	C	A	T	G	C	C	T	G	C	A	C	C	T	G	A	A	C	T	T	C	C	C	A	T	C	A	G	C	C	C	C	A	T	C	G	A	540		
Qy	698	CC	G	T	G	A	G	C	T	G	A	A	C	C	G	C	A	T	G	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	757					
Db	541	CC	G	T	G	A	A	G	C	T	G	A	A	C	C	G	T	G	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	600						
Qy	758	A	G	A	A	T	C	A	A	G	C	C	C	T	A	C	C	C	C	A	T	C	G	A	G	A	G	A	T	G	G	A	A	G	A	G	G	C	A	A	G	T	C	A	C	A	817			
Db	601	A	G	A	A	T	C	A	A	G	C	C	T	T	G	T	G	A	G	A	T	C	G	C	A	C	C	G	A	T	T	G	C	A	C	C	G	A	A	G	A	G	G	C	A	G	A	660		
Qy	818	A	G	A	T	C	G	G	C	C	C	G	A	A	C	C	C	C	T	A	A	C	A	C	C	C	C	C	G	T	T	T	C	G	C	C	A	T	C	A	A	G	A	A	G	A	877			
Db	661	A	G	A	T	C	G	C	C	C	C	G	A	A	C	C	C	C	T	A	A	C	A	C	C	C	C	C	G	T	T	T	C	G	C	C	A	T	C	A	A	G	A	A	G	A	720			
Qy	878	C	C	A	A	G	T	C	G	C	A	A	G	T	T	C	G	C	A	G	A	C	T	T	C	G	C	A	G	A	C	T	T	C	G	A	C	C	C	A	C	C	A	G	A	C	T	937		
Db	721	C	C	A	A	G	T	C																																										

Db	1315	AGATTCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGTCTGCGGGCACCAAGGCC	1374
Qy	1538	TGACCGGACATCGTGCCCTCTACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAACCGCGAGA	1597
Db	1375	TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA	1434
Qy	1598	TCCTGCGGAGCCGCTGCGAGCGGTGTACTACGACCCGAGCAAGGACCTGGTGCCCGAGA	1657
Db	1435	TCCTGGAAGGAGCCGCTGCAAGAGGTGTACTACGACCCGAGCAAGGACCTGGTGCCCGAGA	1484
Qy	1658	TCCAGAAACAGGGCCACGACAGCTGGACCTTACCAGATCTTACCAGAGGCCCTTCAAGAACC	1717
Db	1495	TCCAGAACGAGGGCCAGGCGCCAGTGGACCTTACCAGATCTTACCAGAGGCCCTTCAAGAACC	1554
Qy	1718	TGAAGACGGCAAGTACGCCAGATGCGACCGCCACACCAACGACGTGAAGCAGCTGA	1777
Db	1555	TGAAGACGGCAAGTACGCCCGCATGCGCGCGCCACACCAACGACGTGAAGCAGCTGA	1614
Qy	1778	CCGAGGCGGTGCAAGAATGCCATGAGAGACATCGTGATCTGGGGCAAGACCCCCCAAGT	1837
Db	1615	CCGAGGCGGTGCAAGAAGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT	1674
Qy	1838	TCCGCTTGCCCATCCAGAGGAGACCTTGGAGACCTTGTTGACCGCATCTGGCAGGCCA	1897
Db	1675	TCAAGCTGCCCATCCAGAGGAGACCTTGGAGGCGCTGGTGATGGAGTACTTGGCAGGCCA	1734
Qy	1898	CCTGGATCCCCGAGTGGGAGTTCTGTGAACACCCCCCCCCTGGTGAAGCTGTGTGTAACGAC	1957
Db	1735	CCTGGATCCCCGAGTGGGAGTTCTGTGAACACCCCCCCCCTGGTGAAGCTGTGTGTAACGAC	1794
Qy	1958	TGGAGAGGAGCCCATCATCGCGCCGAGACCTTCTACGTGGAGCGGCGCCGACCAACCGCG	2017
Db	1795	TGGAGAGGAGCCCATCTGTGGCGCCGAGACCTTCTACGTGGAGCGGCGCCGACCAACCGCG	1854
Qy	2018	AGACCAAGATCGGCAAGCCCGGCTACGTGACCGACCGGGCCGGCAGAGAAGATCGTGAGCC	2077
Db	1855	AGACCAAGCTGGGCAAGCCCGGCTACGTGACCGACCGGGCCGGCAGAGAAGTGTGTGAGCA	1914
Qy	2078	TGAACGAGACCAACCAACAGAAACCGAGGTGCAAGGCCATCCAGCTGGCCCTTGACAGACA	2137
Db	1915	TCGCGGACACCAACCAACAGAAACCGAGGTGCAAGGCCATCCACCTGGCCCTTGACAGACA	1974
Qy	2138	CGGCGAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGATCATCCAGGCC	2197
Db	1975	CGGCGCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGATCATCCAGGCC	2034
Qy	2198	AGCCCGCAAGAGCGAGCGAGCTGTGTGAACCAAGATCATCGAGCAGCTGTATCAAGAAGG	2257
Db	2035	AGCCCGCAAGAGCGAGCGAGCTGTGTGAAGCCAGGATACGCCCTGGGATCATCAAGAAGG	2094
Qy	2258	AGAGGTGTACTAGCTGGTGCCGCCCCACCAAGGCGATCGGCGGCAACGAGCAGATCG	2317
Db	2095	AGAGGTGTACTAGCTGGTGCCGCCCCCAAGGCGATCGGCGGCAACGAGCAGATCG	2154
Qy	2318	ACAAGCTGGTGAGCAAGGCCATCCGCAAGGTGTGTTCTTGAGCGGCGATCGATGGCGCA	2377
Db	2155	ACAAGCTGGTGAGCGCGGATCCGCAAGGTGTGTTCTTGAGCGGCGATCGATGGCGCA	2214
Qy	2378	TCGTGATCTACAGTACATGAGACACCTGTACGTGGCAGCGGCGGCCCTTAGGATCGATT	2437
Db	2215	TCGTGATCTACAGTACATGAGACACCTGTACGTGGGCGAGCGGCGGCCCTTAGGATCGATT	2274
Qy	2438	AAAAGCTTCCCGGGCTAGCACCGGTGAATT	2469
Db	2275	AAAAGCTTCCCGGGCTAGCACCGGTGAATT	2306

RESULT 14  
AX427938  
LOCUS  
DEFINITION  
ACCESSION  
VERSION



```
QY 1862 CTTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCC 1921
Db 4818 CTTGGGAGGCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCC 4877
QY 1922 TGAACACCCCCCTGGTGAAGCTGTGGTACAGCTGGAGAGGAGCCATCATCGGCG 1981
Db 4878 TGAACACCCCCCTGGTGAAGCTGTGGTACAGCTGGAGAGGAGCCATCATCGGCG 4937
QY 1982 CCAGAGACCTTCTACGTGGAGCGGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGCGGCT 2041
Db 4938 CCAGAGACCTTCTACGTGGAGCGGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGCGGCT 4997
QY 2042 ACGTGACCCGACCGGGCGCGCAGAGATCGTGAAGCTGTGAGCCGAGACCAACCAAGAGAA 2101
Db 4998 ACGTGACCCGACCGGGCGCGCAGAGATCGTGAAGCTGTGAGCCGAGACCAACCAAGAGAA 5057
QY 2102 CCAGAGCTGAGGCGATCCAGCTGGCCCTGCAGACAGCGGCGAGGCTGAACATCGTGA 2161
Db 5058 CCAGAGCTGAGGCGATCCAGCTGGCCCTGCAGACAGCGGCGAGGCTGAACATCGTGA 5117
QY 2162 CCAGACGCCAGTACGCCCTGGGCATCATCCAGGCCCGACCAAGAGCGAGCGAGC 2221
Db 5118 CCAGACGCCAGTACGCCCTGGGCATCATCCAGGCCCGACCAAGAGCGAGCGAGC 5177
QY 2222 TGGTGAACAGATCATCGACAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGC 2281
Db 5178 TGGTGAACAGATCATCGACAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGC 5237
QY 2282 CCGCCCAACAGAGGATCGGCGGCAACGAGCAGATCGACAGCTGGTGAGCAAGGGCATCC 2341
Db 5238 CCGCCCAACAGAGGATCGGCGGCAACGAGCAGTGGAGCTGGTGAGCGCGGCGATCC 5297
QY 2342 GCAAGTGTCTTCTGGAGCGGATCGATGGCGGCTCGTGATCTACAGTA 2393
Db 5298 GCAAGTGTCTTCTGGAGCGGATCGACAAAGGCCCGAGGAGGACGAGAA 5349

RESULT 15
AX427921
LOCUS AX427921 9189 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 159 from Patent WO0232943.
ACCESSION AX427921
VERSION AX427921.1 GI:21538008
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1
AUTHORS Huang, Y. and Nabel, G. J.
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for
genetic immunization
JOURNAL Patent: WO 0232943-A 159 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES
Location/Qualifiers
1..9189
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"

ORIGIN
Query Match 81.8%; Score 2018.6; DB 6; Length 9189;
Best Local Similarity 91.4%; Pred. No. 5.2e-203;
Matches 2190; Conservative 0; Mismatches 184; Indels 21; Gaps 4;

QY 14 TGGCCGAGGCGCATGAGCCAGG---CCACACGCGCCACATCTGTATGTCAGCGCAGCAACT 70
Db 2967 TGGCCGAGGCGCATGAGCCAGGTTGACCAAGCGGCCACCATCATGATGTCAGCGCGCACT 3026
QY 71 TCAGAGGCGCCCAAGCGCATCATCAAGTGTCTCAACTCGGCGAAGAGGGCCCATCGCCC 130
Db 3027 TCCGCAACCGCGCAAGATCGTGAAGTGTCTCAACTCGGCGAAGAGGGGCCACACGCCCC 3086
```

```
QY 131 GCAACTGCCGCGCCCGCCGCAAGAGAGGCTGTCTGGAAGTCCGGCAAGAGGGGCCACCAGA 190
Db 3087 GCAACTGCCGCGCCCGCCGCAAGAGAGGCTGTCTGGAAGTCCGGCAAGAGGGGCCACCAGA 3146
QY 191 TGAAGGATCGGACCGAGGCGCCAGGCCAACTTTCTTCGCGAGAGACCTTGGCTTCCGCCAGG 250
Db 3147 TGAAGGATCGGACCGAGGCGCCAGGCCAACTTTCTTCGCGAGAGACCTTGGCTTCCGCCAGG 3200
QY 251 GCAAGGCGCCGAGTTCGCCAGCGAGCAGAAACCGCGCCAAACAGCCCCACACAGCCGCGAGC 310
Db 3201 GGAAGGCGAGGGAATTTTCTTCAGAGCAGACAGAGCCAAACAGCCCCACACAGAGAGAGC 3260
QY 311 TGCAGGTGCGCGG-----CGAACAAACCCCGCAGAGCGCGCGCGCGAGCGCCAGGGCA 364
Db 3261 TTCAGGTTTGGGGAAGAGACAAACAACTTCCCTCTCAGAAGCAGAGCGCGATAGACAAAGAA 3320
QY 365 -----CCCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCCCTTGGTGAAGCATCAAGG 418
Db 3321 CTGTATCTTTAGTTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGA 3380
QY 419 TGGCGGCGCAGATCAAGAGGCGCTGTGGACACCGCGCGCGAGCGCGCGAGCACCGTGTGGAG 478
Db 3381 TAGGGGCGCAGCTGAAGAGGCGCTTCTAGACACCGCGCGCGAGCGACACCGTGTGGAGG 3440
QY 479 AGATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTTCATCA 538
Db 3441 AGATGAACCTGCGCGCGCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTTCATCA 3500
QY 539 AGGTGCGCGAGTACGACAGATCTGTGAGATCTGTGGCGCAAGAGGCCATCGGACCG 598
Db 3501 AGGTGGGCGCAGTACGACAGATCTGTGATCGAGATCTGTGGCGCAAGAGGCCATCGGACCG 3560
QY 599 TGTGTATCGGCGCCACCGCGTGAACATCATCGGCGCAACATGCTGACCGAGCTGGGCT 658
Db 3561 TGTGTGTGGGCGCCACCGCGTGAACATCATCGGCGCAACATGCTGACCGAGCTGGGCT 3620
QY 659 GCACCTGAACTTCCCATCAGCCCCATCGAGACCGTGCCTGTAAGTGAAGCCCGGCA 718
Db 3621 GCACCTGAACTTCCCATCAGCCCCATCGAGACCGTGCCTGTAAGTGAAGCCCGGCA 3680
QY 719 TGGAGCGGCGCAAGGTGAGCGCTGCGCCCTGACCGAGGAGAGATCAAGCCCTGACCG 778
Db 3681 TGGACGCGGCGCAAGGTGAAGCGTGGCGCTTACCGAGGAGAGATCAAGGCCCTGTGG 3740
QY 779 CCATCTGCGAGGATGGAAGAGGCGCAAGATCACAAAGATCGGCGCCCGAGAACCCCT 838
Db 3741 AGATCTGACCGAGATGGAAGAGGCGCAAGATCACAAAGATCGGCGCCCGAGAACCCCT 3800
QY 839 ACAACACCCCGCTTTCGCCATCAAGAAAGAGCAGACCAAGTGGCGCAAGCTGGTGG 898
Db 3801 ACAACACCCCGCTTTCGCCATCAAGAAAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGG 3860
QY 899 ACTTCCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCC 958
Db 3861 ACTTCCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCC 3920
QY 959 ACCCGCGCGGCTGAAGAAAGAGCGTGTGACCGCTGTGGAGCTGGGCGAGCGCTTACT 1018
Db 3921 ACCCGCGCGGCTGAAGAAAGAGCGTGTGACCGCTGTGGAGCTGGGCGAGCGCTTACT 3980
QY 1019 TCAGGTGCGCCCTGAGCGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCA 1078
Db 3981 TCAGGTGCGCCCTGAGCGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCA 4040
QY 1079 ACAACGAGACCCCGGCTCCGCTACCAAGTACAAAGTGTGCCCGAGGCTGGAAGGCA 1138
Db 4041 ACAACGAGACCCCGGCTCCGCTACCAAGTACAAAGTGTGCCCGAGGCTGGAAGGCA 4100
QY 1139 GCCCGAGCATCTCCAGAGCAGCATGCCAGATCTCGGCGCCCTCGGCGCCCGCAACC 1198
Db 4101 GCCCGGCGCATCTTCAGTGCAGCATGACCAAGATCTCGGAGCCCTTCGCAAGCAGAC 4160
QY 1199 CCGAGATCGTATCTTACCAGTACATGGACGACCTGTAGTGGGCGAGCACTGGAGATCG 1258
```

Db	4161		CCGACATCGTGATCTACCAAGTACATGACGACCTGTACGTGGCAGGACCTGGAGATCG	4220
Qy	1259		GCCAGCACCGGCGCAAGATCGAGAGCTGGCAAGCACTGCTGCGCTGGGGTTACCA	1318
Db	4221		GCCAGCACCGCACCAAGATCGAGAGCTGGCCAGCACTGCTGCGCTGGGGTTACCA	4280
Qy	1319		CCCCCGCAAGAACACCAAGAGAGCCCGCTTCTGTGGATGGGTACGAGCTGCACC	1378
Db	4281		CCCCCGCAAGAACACCAAGAGAGCCCGCTTCTGTGGATGGGTACGAGCTGCACC	4340
Qy	1379		CCGACAAAGTGACCGTGAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCGTGAACG	1438
Db	4341		CCGACAAAGTGACCGTGAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCGTGAACG	4400
Qy	1439		ACATCCAGAAAGCTGGTGGGCAAGCTGAACCTGGGCGAGCAGATCTACCCCGGATCAAGG	1498
Db	4401		ACATCCAGAAAGCTGGTGGGCAAGCTGAACCTGGGCGAGCAGATCTACCCCGGATCAAGG	4460
Qy	1499		TGCGCCAGCTGTGCAAGCTGTGGCGCGGCCCAAGGCCCTGACCGACATCGTGCCCTGA	1558
Db	4461		TGCGCCAGCTGTGCAAGCTGTGGCGCGGCCCAAGGCCCTGACCGAGTGTGCTCCCTGA	4520
Qy	1559		CCGAGGAGCGGAGCTGGAGCTGGCCGAGAACCGGAGATCCTGGCGAGCCCGTGACG	1618
Db	4521		CCGAGGAGCGGAGCTGGAGCTGGCCGAGAACCGGAGATCCTGAAGAGCCCGTGACG	4580
Qy	1619		GCGTGTACTAGCAACCCAGCAAGACCTGTGTGGCGCGCCAGGCTGACCGAGATCCAGAAGCAGGCGCCACGACC	1678
Db	4581		GCGTGTACTAGCAACCCAGCAAGACCTGTGTGGCGCGCCAGGCTGACCGAGATCCAGAAGCAGGCGCCACGACC	4640
Qy	1679		AGTGAGACTTACAGATCTACAGGAGGCCCTTCAGAACCTGAAGACCGGCAAGTACGCCA	1738
Db	4641		AGTGAGACTTACAGATCTACAGGAGGCCCTTCAGAACCTGAAGACCGGCAAGTACGCCC	4700
Qy	1739		AGATCGGCACCGCCACACCAACGACGTGAAGCAGCTGACCGAGGCGCTGCAGAAAGATCG	1798
Db	4701		GCATGAAGGGCGCCACACCAACGACGTGAAGCAGCTGACCGAGGCGCTGCAGAAAGATCG	4760
Qy	1799		CCATGGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAAAG	1858
Db	4761		CCACCGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAAAG	4820
Qy	1859		AGACCTGGGAGACCTGTGGACCGGCTACTGGCAGGCCACCTGGATCCCGAGTGGGAGT	1918
Db	4821		AGACCTGGGAGGCGCTGTGGACCGGCTACTGGCAGGCCACCTGGATCCCGAGTGGGAGT	4880
Qy	1919		TCGTGAACACCCCGCCCTGGTGAAGCTGTGTATCCAGCTGGAGAGAGGCGCCATCATCG	1978
Db	4881		TCGTGAACACCCCGCCCTGGTGAAGCTGTGTATCCAGCTGGAGAGAGGCGCCATCATCG	4940
Qy	1979		GCGCGAGAGACTTCTACGTGGAGCGGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGG	2038
Db	4941		GCGCGAGAGACTTCTACGTGGAGCGGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGG	5000
Qy	2039		GCTACGTGACCGACCGGCGCGGCAAGATCGTGAGCTGACCGGAGACCAACACCA	2098
Db	5001		GCTACGTGACCGACCGGCGCGGCAAGATCGTGAGCTGACCGGAGACCAACACCA	5060
Qy	2099		AGACCGAGCTGCAGGCGCATCCAGCTGGCCCTGCAGACAGCGGCGAGGTGAACATCG	2158
Db	5061		AGACCGAGCTGCAGGCGCATCCAGCTGGCCCTGCAGACAGCGGCGAGGTGAACATCG	5120
Qy	2159		TGACCGACAGCCAGTACGCGCTGGGCGATCATCAGGCCCGAGCCCGACAGAGCGAGGCG	2218
Db	5121		TGACCGACAGCCAGTACGCGCTGGGCGATCATCAGGCCCGAGCCCGACAGAGCGAGGCG	5180
Qy	2219		AGCTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGG	2278
Db	5181		AGCTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGG	5240
Qy	2279		TGCCCCGCCCAAGAGGCGATCGGCGCAACGAGCAGATCGAACAGCTGTGAGCAAGGGCA	2338

Search completed: April 10, 2004, 12:49:04  
Job time : 6354.08 secs

Db	5241	TGCCCCGCCCAACAAAGGCGATCGGCGGCAACGAGCAGGCTGGTGGACGCGCTGGTGGACCGCGCA	5300
Qy	2339	TCCGCAAGGTGCTGTTCTTCTGGACCGGCATCGATGGCGGCATCGTGTATCTTACCACTA	2393
Db	5301	TCCGCAAGGTGCTGTTCTTCTGGACCGGCATCGACAAAGGCCAGGAGCAGCAGAA	5355

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 02:45:46 ; Search time 622.513 Seconds  
(without alignments)

16849.133 Million cell updates/sec

Title: US-09-610-313-30

Perfect score: 2469  
Sequence: 1 gtgacgccaccatggccga.....gggctagcaccggtgaattc 2469

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001as:\*
- 5: Geneseq2001bs:\*
- 6: Geneseq2002s:\*
- 7: Geneseq2003as:\*
- 8: Geneseq2003bs:\*
- 9: Geneseq2003cs:\*
- 10: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2469	100.0	2469	6	ABL39959
2	2457	99.5	2457	7	ACA03548
3	2457	99.5	2457	9	ADCl3266
4	2442.2	98.9	2463	6	ABL39960
5	2434.6	98.6	2457	7	ACA03547
6	2434.6	98.6	2457	6	ADCl3265
7	2415.4	97.8	2457	6	ABL39961
8	2401.8	97.3	2445	7	ACA03546
9	2401.8	97.3	2445	9	ADCl3264
10	2394.8	97.0	3930	9	ADCl3230
11	2393.2	96.9	3930	9	ADCl3231
12	2393.2	96.9	3930	9	ADCl3232
13	2393.2	96.9	5184	7	ACA03591
14	2393.2	96.9	5184	9	ADCl3279
15	2362.8	95.7	3531	9	ADCl3234
16	2361.2	95.6	3537	9	ADCl3236
17	2360.2	95.6	5145	7	ACA03521
18	2360.2	95.6	5145	9	ADCl3233
19	2350.2	95.2	3538	9	ADCl3235
20	2349.4	95.2	3624	7	ACA03550
21	2349.4	95.2	3624	9	ADCl3268
22	2301.8	93.2	3607	7	ACA03551
23	2301.8	93.2	3607	9	ADCl3269

24 2283.6 92.5 3597 7 ACA03549 Synthetic  
25 2283.6 92.5 3597 9 ADCl3267 DNA of HI  
26 2149.6 87.1 2472 7 ACA03543 Synthetic  
27 2149.6 87.1 2472 7 ACC78507 HIV p2Pol  
28 2121.2 85.9 2466 7 ACA03542 Synthetic  
29 2121.2 85.9 2466 7 ACC78506 HIV p2Pol  
30 2094.4 84.8 2460 7 ACA03541 Synthetic  
31 2094.4 84.8 2460 7 ACC78505 HIV p2Pol  
32 2093.4 84.8 3564 7 ACC78488 HIV GagPo  
33 2093.4 84.8 3564 7 ACC78489 HIV GagPo  
34 2092.8 84.8 4716 7 ACA03522 Synthetic  
35 2092.8 84.8 4716 9 ADCl3238 DNA of HI  
36 2089 84.6 3999 7 ACC78484 HIV GagCo  
37 2087.4 84.5 3999 7 ACC78485 HIV GagCo  
38 2087.4 84.5 3999 7 ACC78486 HIV GagCo  
39 2087.4 84.5 5283 7 ACA03584 Synthetic  
40 2087.4 84.5 5283 7 ACC78529 HIV TatRe  
41 2087.2 84.5 4713 7 ACA03592 Synthetic  
42 2087.2 84.5 4713 9 ADCl3280 DNA of HI  
43 2086.8 84.5 3462 9 ADCl3237 HIV p2Pol  
44 2061.4 83.5 3735 7 ACA03545 Synthetic  
45 2061.4 83.5 3735 7 ACC78509 HIV p2Pol

#### ALIGNMENTS

##### RESULT 1

ABL39959  
ID ABL39959 standard; DNA; 2469 BP.  
XX  
AC ABL39959;  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DE Synthetic construct PR975(+) SEQ ID NO:30.  
XX  
KW Human immunodeficiency virus type C; antigenic HIV type C protein;  
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;  
KW immunostimulant; gene therapy; gene; ds.  
XX  
OS Human immunodeficiency virus; type C.  
OS Synthetic.  
XX  
FN WO200204493-A2.  
XX  
PD 17-JAN-2002.  
XX  
PF 05-JUL-2001; 2001WO-US021241.  
XX  
PR 05-JUL-2000; 2000US-00610313.  
XX  
PA (CHIR ) CHIRON CORP.  
XX (UYST-) UNIV STELLENBOSCH.  
XX  
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;  
XX  
PS WPI; 2002-154920/20.  
XX  
FT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in  
FT applications including DNA immunization or generation of packaging cell  
FT lines, particularly in gene therapy.

Claim 1; Fig 8; 233pp; English.

The present invention describes expression cassettes comprising a polynucleotide sequence encoding a polypeptide comprising immunogenic HIV type C polypeptides. The expression cassettes comprise any of the HIV type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef (i). (i) have immunostimulant activity and can be used in gene therapy. The HIV type C polynucleotides are useful in applications including DNA immunisation, generation of packaging cell lines, and production of HIV Type C proteins. The polynucleotides are particularly useful in gene

CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
CC ABB06204 to ABB06215 represent sequences used in the exemplification of  
CC the present invention  
XX  
SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;  
  
Query Match 100.0%; Score 2469; DB 6; Length 2469;  
Best Local Similarity 100.0%; Pred. No. 5.2e-298;  
Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GTGACGCCACCATGGCCGAGGCCATGAGCAGGCCACAGCCCAACATCTCTGATGCAG 60  
DB 1 GTGACGCCACCATGGCCGAGGCCATGAGCAGGCCACAGCCCAACATCTCTGATGCAG 60  
  
QY 61 CGCAGCAACTTCAAGGGCCCAAGCCGATCATCAAGTGTCTTCACTGCGGCAAGGAGGCG 120  
DB 61 CGCAGCAACTTCAAGGGCCCAAGCCGATCATCAAGTGTCTTCACTGCGGCAAGGAGGCG 120  
  
QY 121 CACATCCCGCCCAACTCCCGCGCCCGCCGCAAGAGGGCTGTCTGGAAGTGCAGCAAGGAG 180  
DB 121 CACATCCCGCCCAACTCCCGCGCCCGCCGCAAGAGGGCTGTCTGGAAGTGCAGCAAGGAG 180  
  
QY 181 GGCACACAGATGAAGACTGACCGAGCGCCAGCCCAACTTCTTCGCGAGGACCTGGCC 240  
DB 181 GGCACACAGATGAAGACTGACCGAGCGCCAGCCCAACTTCTTCGCGAGGACCTGGCC 240  
  
QY 241 TTCCCGCAGGCAAGGCCCGCGAGTTCCCGAGCGAGCAGAACCGCGCAACAGCCCGCAC 300  
DB 241 TTCCCGCAGGCAAGGCCCGCGAGTTCCCGAGCGAGCAGAACCGCGCAACAGCCCGCAC 300  
  
QY 301 AGCGCGAGCTGACAGTGCAGGCGGCGCAACACCGCGAGCGAGCGCGCGCGAGCGCCAG 360  
DB 301 AGCGCGAGCTGACAGTGCAGGCGGCGCAACACCGCGAGCGAGCGCGCGCGAGCGCCAG 360  
  
QY 361 GGCACTCTGAATTCACAGAGGCGCTGCTGGAACCGCGCGAGCAGACCGTGTCTGAGGAG 420  
DB 361 GGCACTCTGAATTCACAGAGGCGCTGCTGGAACCGCGCGAGCAGACCGTGTCTGAGGAG 420  
  
QY 421 GGCGGCCAGATCAAGAGGCGCTGCTGGAACCGCGCGAGCAGACCGTGTCTGAGGAG 480  
DB 421 GGCGGCCAGATCAAGAGGCGCTGCTGGAACCGCGCGAGCAGACCGTGTCTGAGGAG 480  
  
QY 481 ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAG 540  
DB 481 ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAG 540  
  
QY 541 GTGCGCCAGTACGACACAGATCTGATCGAGATCTGCGGCAAGAGGCCATCGGCAACG 600  
DB 541 GTGCGCCAGTACGACACAGATCTGATCGAGATCTGCGGCAAGAGGCCATCGGCAACG 600  
  
QY 601 CTGATCGGCCCGACCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
DB 601 CTGATCGGCCCGACCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
  
QY 661 ACCCTGAACTTCCCATCAGCCCATCGAGACCGTGCCTGAGCTGAAGCCCGGCATG 720  
DB 661 ACCCTGAACTTCCCATCAGCCCATCGAGACCGTGCCTGAGCTGAAGCCCGGCATG 720  
  
QY 721 GACGCCCGCAAGTGAAGCAGTGGCCCTGACCGAGGAGATCAAGGCGCTGACCGCC 780  
DB 721 GACGCCCGCAAGTGAAGCAGTGGCCCTGACCGAGGAGATCAAGGCGCTGACCGCC 780  
  
QY 781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCAACAGATCGGCCCGCGAGACCCCTAC 840  
DB 781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCAACAGATCGGCCCGCGAGACCCCTAC 840  
  
QY 841 AACACCCCGCTTTCGCCATCAAGAGAGGAGCAGCAGCAAGTGGCGCAAGCTGGTGGAC 900  
DB 841 AACACCCCGCTTTCGCCATCAAGAGAGGAGCAGCAGCAGTGGCGCAAGCTGGTGGAC 900  
  
QY 901 TTCCCGGAGCTGAAACAAGCGCACCCAGACTTCTGGGAGGTGCGAGTGGGCATCCCCAC 960  
DB 901 TTCCCGGAGCTGAAACAAGCGCACCCAGACTTCTGGGAGGTGCGAGTGGGCATCCCCAC 960

QY 961 CCCCGCGCTTGAAGAAGAAAGAGCGTGAACCGTCTGAGCGTGGCGGAGCGCTACTTC 1020  
DB 961 CCCCGCGCTTGAAGAAGAAAGAGCGTGAACCGTCTGAGCGTGGCGGAGCGCTACTTC 1020  
  
QY 1021 AGCGTCCCGCTGGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGCAGCATCAAC 1080  
DB 1021 AGCGTCCCGCTGGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGCAGCATCAAC 1080  
  
QY 1081 AACGAGACCCCGGCATCCGCTACAGTAAAGTGTGCTGCCAGGCTGGAGGGGAGC 1140  
DB 1081 AACGAGACCCCGGCATCCGCTACAGTAAAGTGTGCTGCCAGGCTGGAGGGGAGC 1140  
  
QY 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGAGCCCTTCCGCGCCGCAACCCC 1200  
DB 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGAGCCCTTCCGCGCCGCAACCCC 1200  
  
QY 1201 GAGATCGTGTACTACAGTACATGGACGACCTGTGCTGCGCTGGGCTTCAACACC 1260  
DB 1201 GAGATCGTGTACTACAGTACATGGACGACCTGTGCTGCGCTGGGCTTCAACACC 1260  
  
QY 1261 CAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACTGTGCTGCGCTGGGCTTCAACACC 1320  
DB 1261 CAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACTGTGCTGCGCTGGGCTTCAACACC 1320  
  
QY 1321 CCCGCAAGAAAGCACCAGAGAGGCCCTTCTGTGAGTGGGCTACGAGCTGCAACCCC 1380  
DB 1321 CCCGCAAGAAAGCACCAGAGAGGCCCTTCTGTGAGTGGGCTACGAGCTGCAACCCC 1380  
  
QY 1381 GACAGTGGACCGTGCAGCCATCGAGCTGCCGAGAGGAGAGCTGAGCCGTGAAGCAG 1440  
DB 1381 GACAGTGGACCGTGCAGCCATCGAGCTGCCGAGAGGAGAGCTGAGCCGTGAAGCAG 1440  
  
QY 1441 ATCCAGAAAGCTGGTGGGCAAGCTGAACTGGGCCAGGAGATCTACCCCGGATCAAGGTG 1500  
DB 1441 ATCCAGAAAGCTGGTGGGCAAGCTGAACTGGGCCAGGAGATCTACCCCGGATCAAGGTG 1500  
  
QY 1501 CGCAGCTGTGCAAGCTGCTGCGCGCGCCAGGCGCTGACCGACATCTGCCCCCTGACC 1560  
DB 1501 CGCAGCTGTGCAAGCTGCTGCGCGCGCCAGGCGCTGACCGACATCTGCCCCCTGACC 1560  
  
QY 1561 GAGGAGGCGAGCTGGAGCTGCGCGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCACGGC 1620  
DB 1561 GAGGAGGCGAGCTGGAGCTGCGCGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCACGGC 1620  
  
QY 1621 GTGTACTACAGCCCGCAGCAAGACCTGCTGCGCGAGATCCAGAGAGCGGCGCAGCAG 1680  
DB 1621 GTGTACTACAGCCCGCAGCAAGACCTGCTGCGCGAGATCCAGAGAGCGGCGCAGCAG 1680  
  
QY 1681 TGGACCTACAGATCTACAGAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAG 1740  
DB 1681 TGGACCTACAGATCTACAGAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAG 1740  
  
QY 1741 ATGCGCACCGGCCACCAACAGAGTGAAGCAGCTGACCGAGGCGGCTGCAAGATCGCC 1800  
DB 1741 ATGCGCACCGGCCACCAACAGAGTGAAGCAGCTGACCGAGGCGGCTGCAAGATCGCC 1800  
  
QY 1801 ATGAGAGCATCTGATCTGCGGCAAGACCCCGCAAGTTCCCGCTGCCCATCCAGAAGGAG 1860  
DB 1801 ATGAGAGCATCTGATCTGCGGCAAGACCCCGCAAGTTCCCGCTGCCCATCCAGAAGGAG 1860  
  
QY 1861 ACCTGGGAGCATGCTGTGAGCAGATCTGAGGAGCGACCTGATCCCGGAGTGGAGTTC 1920  
DB 1861 ACCTGGGAGCATGCTGTGAGCAGATCTGAGGAGCGACCTGATCCCGGAGTGGAGTTC 1920  
  
QY 1921 GTGAACACCCCGCTTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980  
DB 1921 GTGAACACCCCGCTTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980  
  
QY 1981 GCCGAGACCTTCTACGTGAGCGGCCCGCAACCGCGAGACCAAGATCGGCAAGGCCCGGC 2040  
DB 1981 GCCGAGACCTTCTACGTGAGCGGCCCGCAACCGCGAGACCAAGATCGGCAAGGCCCGGC 2040





```
Db 781 GAGAGATGAGAAAGGAGGCAAGATCAACAGATCGGCCCGGAGAACCCCTACACACC 840
Qy CCGGTGTTCCGCAATCAAGAAAGAGGACAGACCAAGTGGGCGAAGCTGTGTGATTTCCGC 906
Db 841 CCGGTGTTCCGCAATCAAGAAAGAGGACAGACCAAGTGGGCGAAGCTGTGTGATTTCCGC 900
Qy GAGCTGAACAGCGCACCCAGGACTTCTGGAGAGTGCAGCTGGCATCCGCCACCCCGCC 966
Db 901 GAGCTGAACAGCGCACCCAGGACTTCTGGAGAGTGCAGCTGGCATCCGCCACCCCGCC 960
Qy GGCTGAAGAAAGAAAGAGCGTGAACCGTCTGGAGCGTGGCGACCGCTACTTTACCGTG 1026
Db 961 GGCTGAAGAAAGAAAGAGCGTGAACCGTCTGGAGCGTGGCGACCGCTACTTTACCGTG 1020
Qy CCCCTGACAGGAGTTCGCCAGTACACCGCTTACACCATCCCGACGATCAACACAGAG 1086
Db 1021 CCCCTGACAGGAGTTCGCCAGTACACCGCTTACACCATCCCGACGATCAACACAGAG 1080
Qy ACCCCCGGCATCCGCTACCAAGTACAACGTGTGTCGCCCGAGGGCTGGAAAGGCGACCCCAGC 1146
Db 1081 ACCCCCGGCATCCGCTACCAAGTACAACGTGTGTCGCCCGAGGGCTGGAAAGGCGACCCCAGC 1140
Qy ATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGCCCGCAACCCCGGATC 1206
Db 1141 ATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGCCCGCAACCCCGGATC 1200
Qy GTGATCTACCAAGTACATGGAGCACTGTAGCTGGGACGACCTGGAGATCGGCCAGCAC 1266
Db 1201 GTGATCTACCAAGTACATGGAGCACTGTAGCTGGGACGACCTGGAGATCGGCCAGCAC 1260
Qy CGGCCAAGATCGAGGAGCTGCGAAGCATCTGTGCTGGCTGGGGCTTACCAACCCCGGAC 1326
Db 1261 CGGCCAAGATCGAGGAGCTGCGAAGCATCTGTGCTGGCTGGGGCTTACCAACCCCGGAC 1320
Qy AAGAAGCACCAAGAGGAGCCCGCTTCTGTGGATGGCTACGAGCTGCACCCCGCAAG 1386
Db 1321 AAGAAGCACCAAGAGGAGCCCGCTTCTGTGGATGGCTACGAGCTGCACCCCGCAAG 1380
Qy TGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGAGCCGTGAAACGACATCCAG 1446
Db 1381 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGAGCCGTGAAACGACATCCAG 1440
Qy AAGCTGTGGGCAAGCTGAACTGGGGCCAGCAGATCTACCCCGCATCAAGGTGGCGCAG 1506
Db 1441 AAGCTGTGGGCAAGCTGAACTGGGGCCAGCAGATCTACCCCGCATCAAGGTGGCGCAG 1500
Qy CTGTGCAGAGCTGTGGCGCGCAAGGCCCTGACCGACATCTGTGGCCCTTGACCGAGGAG 1566
Db 1501 CTGTGCAGAGCTGTGGCGCGCAAGGCCCTGACCGACATCTGTGGCCCTTGACCGAGGAG 1560
Qy CCGAGCTGGAGCTGGCGGAGAACCGGAGATCTCTGGCGAGCCCGTGACCGCGGTGTAC 1626
Db 1561 CCGAGCTGGAGCTGGCGGAGAACCGGAGATCTCTGGCGAGCCCGTGACCGCGGTGTAC 1620
Qy TACGACCCAGCAAGGACCTGTGTGGCGAGATCCAGAAAGAGGCGCCACGACCAAGTGGAC 1686
Db 1621 TACGACCCAGCAAGGACCTGTGTGGCGAGATCCAGAAAGAGGCGCCACGACCAAGTGGAC 1680
Qy TACCAGATCTACAGGAGCCCTTCAAGAACCTGAGACCCGCAAGTACGCCAGATCGGC 1746
Db 1681 TACCAGATCTACAGGAGCCCTTCAAGAACCTGAGACCCGCAAGTACGCCAGATCGGC 1740
Qy ACCGCCACCAACAGAGCTGAAGCAGCTGACCGAGGCGTGGCAGAAAGATCGCCATGGAG 1806
Db 1741 ACCGCCACCAACAGAGCTGAAGCAGCTGACCGAGGCGTGGCAGAAAGATCGCCATGGAG 1800
Qy AGCATCTGTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGG 1866
Db 1801 AGCATCTGTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGG 1860
Qy GAGACCTGTGGACGACTACTTGGCAGGCCACCTGGATCCCGGAGTGGAGTTCTGTGAAC 1926
```

```
Db 1861 GAGACCTGTGGACCGGACTACTGGCAGGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAAC 1920
Qy 1927 ACCCCCGCTGTGTGAGCTGTGTACAGCTGGAGAAGGAGCCCATCATCGCGCGCGAG 1986
Db 1921 ACCCCCGCTGTGTGAGCTGTGTACAGCTGGAGAAGGAGCCCATCATCGCGCGCGAG 1980
Qy 1987 ACCTTTACGTGTGAGCGGCGCGCCCAACCGGAGACCAAGATCGGCAAGGCGGCTACGTG 2046
Db 1981 ACCTTTACGTGTGAGCGGCGCGCCCAACCGGAGACCAAGATCGGCAAGGCGGCTACGTG 2040
Qy 2047 ACCGACCGGGCGCGGAGAGATCTGTAGCTGACCGAGACCCACCAACCAAGAGACCGAG 2106
Db 2041 ACCGACCGGGCGCGGAGAGATCTGTAGCTGACCGAGACCCACCAACCAAGAGACCGAG 2100
Qy 2107 CTGACAGGCGCATCTCCAGCTGGCCCTGCAGGACAGCGGCGAGGAGTGAACATCTGTGAACCGAC 2166
Db 2101 CTGACAGGCGCATCTCCAGCTGGCCCTGCAGGACAGCGGCGAGGAGTGAACATCTGTGAACCGAC 2160
Qy 2167 AGCCAGTACGCTTGGGCGATCATCCAGGCCCGGACGAGCGGAGCGAGCGAGCTGCTG 2226
Db 2161 AGCCAGTACGCTTGGGCGATCATCCAGGCCCGGACGAGCGGAGCGAGCGAGCTGCTG 2220
Qy 2227 AACAGATCATCGAGCAGCTGATCAAGAAAGGAGAGGAGTGTACCTGAGCTGGGTGCCCGCC 2286
Db 2221 AACAGATCATCGAGCAGCTGATCAAGAAAGGAGAGGAGTGTACCTGAGCTGGGTGCCCGCC 2280
Qy 2287 CACAAGGCGATCGGCGGCAACGAGCAGATCGAAGCTGTGTGAGCAGGAGGAGGATCCGCAAG 2346
Db 2281 CACAAGGCGATCGGCGGCAACGAGCAGATCGAAGCTGTGTGAGCAGGAGGAGGATCCGCAAG 2340
Qy 2347 GTGCTGTTCTTGGACGGCATCGATGGGGCATCTCGTGATCTACCAAGTACATGGAGACCTG 2406
Db 2341 GTGCTGTTCTTGGACGGCATCGATGGGGCATCTCGTGATCTACCAAGTACATGGAGACCTG 2400
Qy 2407 TACGTGGCAGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 2463
Db 2401 TACGTGGCAGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 2457
```

## RESULT 3

ADCI3266  
ID ADCI3266 standard; DNA; 2457 BP.

AC ADCI3266;

XX 18-DEC-2003 (first entry)

DE DNA of HIV construct p2Pol-opt\_C SEQ ID NO 45.

XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.

XX Human immunodeficiency virus.

XX WO2003004620-A2.

XX 16-JAN-2003.

XX 05-JUL-2002; 2002WO-US021420.

XX 05-JUL-2001; 2001US-0303192P.

XX 31-AUG-2001; 2001US-031866P.

XX 16-JAN-2002; 2002US-0349871P.

XX (CHIR ) CHIRON CORP.

XX (UYST-) UNIV STELLENBOSCH.

XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg BJ;

XX WPI; 2003-221593/21.

XX New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,

PT Prot, or Rev polypeptide, useful for immunization, or generating  
 PT packaging cell lines.  
 XX Disclosure; Fig 42; 301pp; English.  
 XX The invention relates to a novel expression cassette comprising a  
 CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
 CC Int, Nef, p15RaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
 CC expression cassette can be used to treat HIV type C by gene therapy or  
 CC used in the development of a vaccine. The gene delivery vector is  
 CC administered intramuscularly, intramuscularly, intranasally,  
 CC subcutaneously, intradermally, transdermally, intravaginally,  
 CC intrarectally, orally or intravenously. The expression cassette is useful  
 CC for immunisation, generating packaging cell lines and producing HIV  
 CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
 CC Type C related sequence of the invention.  
 XX  
 SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;  
 Query Match 99.5%; Score 2457; DB 9; Length 2457;  
 Best Local Similarity 100.0%; Pref. No. 1.6e-256;  
 Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 GCCACCATGGCGAGGCGCATGAGCGCCACCGAGCGCCAAACATCTTGATCGACGCGAGC 66  
 DB 1 GCCACCATGGCGAGGCGCATGAGCGCCACCGAGCGCCAAACATCTTGATCGACGCGAGC 60  
 QY 67 AACTTCAGAGGCGCCAGCGGCGATCATCAAGTGTCTCACTCGGCGAGGCGCCACATC 126  
 DB 61 AACTTCAGAGGCGCCAGCGGCGATCATCAAGTGTCTCACTCGGCGAGGCGCCACATC 120  
 QY 127 GCCCGCAACTGCGCGCCCGCCCGCAAGAGGCGTCTGGAAGTGGCGCAAGAGGCGCCAC 186  
 DB 121 GCCCGCAACTGCGCGCCCGCCCGCAAGAGGCGTCTGGAAGTGGCGCAAGAGGCGCCAC 180  
 QY 187 CAGATGAGAGTGTACCGAGCGCGCGAGCGCGCAACTTCTTCGCGAGGAGCTGGCTTCCCC 246  
 DB 181 CAGATGAGAGTGTACCGAGCGCGCGAGCGCGCAACTTCTTCGCGAGGAGCTGGCTTCCCC 240  
 QY 247 CAGGCGCAAGGCGCGGAGTTCCTCCAGCGAGAGAGACCGCGCCAAACAGCGCCACCGCGC 306  
 DB 241 CAGGCGCAAGGCGCGGAGTTCCTCCAGCGAGAGAGACCGCGCCAAACAGCGCCACCGCGC 300  
 QY 307 GAGCTGCGAGTGGCGGCGAGCAACCCCGCGAGCGAGGCGCGCGCGCGCGCGCGCGCGCGC 366  
 DB 301 GAGCTGCGAGTGGCGGCGAGCAACCCCGCGAGCGAGGCGCGCGCGCGCGCGCGCGCGC 360  
 QY 367 CTGAACTTCCCGAGATCACTGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 426  
 DB 361 CTGAACTTCCCGAGATCACTGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 420  
 QY 427 CAGATCAAGAGGCGCGTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 486  
 DB 421 CAGATCAAGAGGCGCGTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 480  
 QY 487 CTGCGCGGCAAGTGGAGC 546  
 DB 481 CTGCGCGGCAAGTGGAGC 540  
 QY 547 CAGTACGACCGATCTGATCGAGATCTGCGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGC 606  
 DB 541 CAGTACGACCGATCTGATCGAGATCTGCGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGC 600  
 QY 607 GCGCCCGACCGCGTGAACATCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 666  
 DB 601 GCGCCCGACCGCGTGAACATCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 660  
 QY 667 AACTTCCCGCATGAGC 726  
 DB 661 AACTTCCCGCATGAGC 720  
 QY 727 CCCAAGGTGAAGCGTGGC 786

DB 721 CCCAAGGTGAAGCGTGGC 780  
 QY 787 GAGGAGATGGAGAGGAGGCGAGATCAACAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 846  
 DB 781 GAGGAGATGGAGAGGAGGCGAGATCAACAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 840  
 QY 847 CCGGTGTTCCCATCAAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 906  
 DB 841 CCGGTGTTCCCATCAAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 900  
 QY 907 GAGCTGAACAAACGCGCACCGAGGACTTCTGGAGGTGCGAGCTGGGGATCCCCACCGCGCGC 966  
 DB 901 GAGCTGAACAAACGCGCACCGAGGACTTCTGGAGGTGCGAGCTGGGGATCCCCACCGCGCGC 960  
 QY 967 GCGCTGAAGAAAGC 1026  
 DB 961 GCGCTGAAGAAAGC 1020  
 QY 1027 CCGCTGGACGAGGAGCTTCCGCAAGTACACCGCGCTTACCATCCCCCAGAGATCAACAGAGAG 1086  
 DB 1021 CCGCTGGACGAGGAGCTTCCGCAAGTACACCGCGCTTACCATCCCCCAGAGATCAACAGAGAG 1080  
 QY 1087 ACCCGCGGATCCGCTACCAAGTACACCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1146  
 DB 1081 ACCCGCGGATCCGCTACCAAGTACACCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1140  
 QY 1147 ATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCGCTTCCGCGCGCGCGCAACCGCGAGATC 1206  
 DB 1141 ATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCGCTTCCGCGCGCGCGCAACCGCGAGATC 1200  
 QY 1207 GTGATCTACAGTACATGAGAGCGACTGTGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGC 1266  
 DB 1201 GTGATCTACAGTACATGAGAGCGACTGTGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGC 1260  
 QY 1267 CGCGCGCAAGATCGAGAGGCTGGCGCAAGCACTGCTGCGTGGGGTTCACCAACCGCGAGC 1326  
 DB 1261 CGCGCGCAAGATCGAGAGGCTGGCGCAAGCACTGCTGCGTGGGGTTCACCAACCGCGAGC 1320  
 QY 1327 AGAAGCAACAGAGAGC 1386  
 DB 1321 AGAAGCAACAGAGAGC 1380  
 QY 1387 TGACACGTCGAGC 1446  
 DB 1381 TGACACGTCGAGC 1440  
 QY 1447 AAGCTGCTGGCGCAAGTGAACCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1506  
 DB 1441 AAGCTGCTGGCGCAAGTGAACCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1500  
 QY 1507 CTGTGCAAGTCTGC 1566  
 DB 1501 CTGTGCAAGTCTGC 1560  
 QY 1567 GCGAGCTGGAGCTGGC 1626  
 DB 1561 GCGAGCTGGAGCTGGC 1620  
 QY 1627 TAGAAGCGCGAGC 1686  
 DB 1621 TAGAAGCGCGAGC 1680  
 QY 1687 TACAGATCTACAGAGAGCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 1746  
 DB 1681 TACAGATCTACAGAGAGCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 1740  
 QY 1747 ACCGCGCGAGC 1806  
 DB 1741 ACCGCGCGAGC 1800  
 QY 1807 AGCATCTGTGATCTGGGGCAAGC 1866  
 DB 1801 AGCATCTGTGATCTGGGGCAAGC 1860

1867 GAGACCTGCTGACCGACTACTGCGAGCCACCTCGATCCCGAGTGGAGTTGTAAC 1926  
 1861 GAGACCTGCTGACCGACTACTGCGAGCCACCTCGATCCCGAGTGGAGTTGTAAC 1920  
 1927 ACCCCCCCTCTGTAAGCTGTGTGACCAAGTGGAGAGGAGCCCATCATCGGCGCCGAG 1986  
 1921 ACCCCCCCTCTGTAAGCTGTGTGACCAAGTGGAGAGGAGCCCATCATCGGCGCCGAG 1980  
 1987 ACTTCTAGCTGACCGCGCCCGCCACCGCGAGACCAAGATCGGCAAGCGCGTACCTG 2046  
 1981 ACTTCTAGCTGACCGCGCCCGCCACCGCGAGACCAAGATCGGCAAGCGCGTACCTG 2040  
 2047 ACCGACCGCGCGCGCGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAGACCCGAG 2106  
 2041 ACCGACCGCGCGCGCGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAGACCCGAG 2100  
 2107 CTGACGGCCATCCAGCTGCGCTGCGAGACAGCGCGAGGAGTGAACATGTGACCGAC 2166  
 2101 CTGACGGCCATCCAGCTGCGCTGCGAGACAGCGCGAGGAGTGAACATGTGACCGAC 2160  
 2167 AGCCAGTACGCTGCGCTGCGCATCTCCAGGCGCCCGCCGACCAAGAGCGAGAGCTGGTG 2226  
 2161 AGCCAGTACGCTGCGCTGCGCATCTCCAGGCGCCCGCCGACCAAGAGCGAGAGCTGGTG 2220  
 2227 AACGAGATCATCGAGAGCTGATCAAGAGAGAGAGTGTACTGAGCTGGTGCGCCGCG 2286  
 2221 AACGAGATCATCGAGAGCTGATCAAGAGAGAGAGTGTACTGAGCTGGTGCGCCGCG 2280  
 2287 CACAAGGATCGCGCGCGCAACAGAGATCGCAAGCTGGTGAAGAGGATCCGCGAAG 2346  
 2281 CACAAGGATCGCGCGCGCAACAGAGATCGCAAGCTGGTGAAGAGGATCCGCGAAG 2340  
 2347 GTCTGTCTTCTGACCGCATCGATGCGCGAGTGTGATCTACAGTACATGACGACCTG 2406  
 2341 GTCTGTCTTCTGACCGCATCGATGCGCGAGTGTGATCTACAGTACATGACGACCTG 2400  
 2407 TACGTGGGAGCGCGCGCTGATGATCGATTAAGCTTCCCGGCTAGCACCGGT 2463  
 2401 TACGTGGGAGCGCGCGCTGATGATCGATTAAGCTTCCCGGCTAGCACCGGT 2457

RESULT 4  
 ABL39960  
 ID ABL39960 standard; DNA; 2463 BP.  
 XX AC ABL39960;  
 XX DT 15-MAY-2002 (first entry)  
 XX DE Synthetic construct PR975YM SEQ ID NO:31.  
 XX KW Human immunodeficiency virus type C; antigenic HIV type C protein;  
 XX KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;  
 XX KW immunostimulant; gene therapy; gene; ds.  
 XX OS Human immunodeficiency virus; type C.  
 XX OS Synthetic.  
 XX PN WO200204493-A2.  
 XX PD 17-JAN-2002.  
 XX PF 05-JUL-2001; 2001WO-US021241.  
 XX PR 05-JUL-2000; 2000US-00610313.  
 XX PA (CHIR ) CHIRON CORP.  
 XX PA (UYST-) UNIV STELLENBOSCH.  
 XX PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg E;  
 XX WI; 2002-154920/20.

XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in  
 PT applications including DNA immunization or generation of packaging cell  
 PT lines, particularly in gene therapy.  
 XX Claim 1; Fig 9; 233pp; English.  
 XX The present invention describes expression cassettes comprising a  
 CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV  
 CC type C polypeptides. The expression cassettes comprise any of the HIV  
 CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef  
 CC (1). (1) have immunostimulant activity and can be used in gene therapy.  
 CC The HIV type C polynucleotides are useful in applications including DNA  
 CC immunisation, generation of packaging cell lines, and production of HIV  
 CC type C proteins. The polynucleotides are particularly useful in gene  
 CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
 CC ABL06204 to ABL06215 represent sequences used in the exemplification of  
 CC the present invention  
 XX Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;  
 SQ Query Match 98.9%; Score 2442.2; DB 6; Length 2463;  
 Best Local Similarity 99.6%; Pred. No. 1.1e-294;  
 Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;  
 QY 1 GTGACGCCACCATGCGCGAGCGCATGAGCCAGGCCACCGAGCCCAACATCTCTGATGCGAG 60  
 DB 1 GTGACGCCACCATGCGCGAGCGCATGAGCCAGGCCACCGAGCCCAACATCTCTGATGCGAG 60  
 QY 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGC 120  
 DB 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGC 120  
 QY 121 CACATCGCCCGCAACTGCGCGCGCCCGCCGCAAGAGGGTCTGGAAGTTCGCGCAAGAGGAG 180  
 DB 121 CACATCGCCCGCAACTGCGCGCGCCCGCCGCAAGAGGGTCTGGAAGTTCGCGCAAGAGGAG 180  
 QY 181 GGCACACAGATGAAGGACTGACCGAGCGCCAGCGCAACTTCTTCCGCGAGGACCTTGGCC 240  
 DB 181 GGCACACAGATGAAGGACTGACCGAGCGCCAGCGCAACTTCTTCCGCGAGGACCTTGGCC 240  
 QY 241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCTCCAGCGAGCGAGAACCGCGCCACACCCGCCACC 300  
 DB 241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCTCCAGCGAGCGAGAACCGCGCCACACCCGCCACC 300  
 QY 301 AGCGCGAGTTCAGAGTTCGCGCGCGCAACACCCCGCGAGCGCGCGCGCGCGCGCGCGAG 360  
 DB 301 AGCGCGAGTTCAGAGTTCGCGCGCGCAACACCCCGCGAGCGCGCGCGCGCGCGCGCGAG 360  
 QY 361 GGCACCTGTGAATTCCTCCCGAGTACCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGAGTG 420  
 DB 361 GGCACCTGTGAATTCCTCCCGAGTACCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGAGTG 420  
 QY 421 GCGCGCGAGATCAAGAGAGGCGCTGCTGGAACACCGCGCGCGCGCGCGCGCGCGCGCGAG 480  
 DB 421 GCGCGCGAGATCAAGAGAGGCGCTGCTGGAACACCGCGCGCGCGCGCGCGCGCGCGAG 480  
 QY 481 ATGAGCTTCCCGCGCAAGTGAAGCGCGCAAGATGATGCGCGCGCGCGCGCGCGCGCGCGAG 540  
 DB 481 ATGAGCTTCCCGCGCAAGTGAAGCGCGCAAGATGATGCGCGCGCGCGCGCGCGCGCGAG 540  
 QY 541 GTGCGCGAGTACGACAGATCTGTGAGATCTGTGCGGCAAGAGGCGCATCGGCGCGCGTG 600  
 DB 541 GTGCGCGAGTACGACAGATCTGTGAGATCTGTGCGGCAAGAGGCGCATCGGCGCGCGTG 600  
 QY 601 CTGATCGGCGAGTGGGTCG 660  
 DB 601 CTGATCGGCGAGTGGGTCG 660  
 QY 661 ACCCTGAGCTTCCCGCGCAAGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGT 720  
 DB 661 ACCCTGAGCTTCCCGCGCAAGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGT 720





QY 1681 TGGACCTACAGATCTACAGGAGCCCTTCAAGAACTGAGACCGGCGAGTACGCCAAG 1740  
Db 1675 TGGACCTACAGATCTACAGGAGCCCTTCAAGAACTGAGACCGGCGAGTACGCCAAG 1734  
QY 1741 ATGCGACCGCCACACCAACGACGCTGAAGACGCTGACGAGGCGCGTGCAGAGATCGCC 1800  
Db 1735 ATGCGACCGCCACACCAACGACGCTGAAGACGCTGACGAGGCGCGTGCAGAGATCGCC 1794  
QY 1801 ATGGAGAGCATGCTGATCTGGGCGAGACCCCAAGTTCCGCTGCCATCCAGAGAGAG 1860  
Db 1795 ATGGAGAGCATGCTGATCTGGGCGAGACCCCAAGTTCCGCTGCCATCCAGAGAGAG 1854  
QY 1861 ACCTGGAGAGCCTGGTGGACCGCACTACTGGCAGCCACCTGGATCCCGAGTGGGATTC 1920  
Db 1855 ACCTGGAGAGCCTGGTGGACCGCACTACTGGCAGCCACCTGGATCCCGAGTGGGATTC 1914  
QY 1921 GTGAACACCCCGCCCTGCTGAGCTGTGTACAGCTGAGAGAGGCCCATCATCGGC 1980  
Db 1915 GTGAACACCCCGCCCTGCTGAGCTGTGTACAGCTGAGAGAGGCCCATCATCGGC 1974  
QY 1981 GCCGAGACCTTCTACGTGGACGGCGCCGCAACCGCAGACCAAGATCGGCAAGGCGGC 2040  
Db 1975 GCCGAGACCTTCTACGTGGACGGCGCCGCAACCGCAGACCAAGATCGGCAAGGCGGC 2034  
QY 2041 TACGTGACCGACCGCGCGCGGAGAGATCGTGAAGCTGACCGAGACCAACACCGAAG 2100  
Db 2035 TACGTGACCGACCGCGCGCGGAGAGATCGTGAAGCTGACCGAGACCAACACCGAAG 2094  
QY 2101 ACCGAGCTGACAGGCGCATCCAGCTGGCCCTGAGCAGCAGCGCAGGAGTGAACATCGT 2160  
Db 2095 ACCGAGCTGACAGGCGCATCCAGCTGGCCCTGAGCAGCAGCGCAGGAGTGAACATCGT 2154  
QY 2161 ACCGAGCGCCAGTACGCGCTGGGCATCATCCAGCGCCAGCCGACAGAGCGAGCGAG 2220  
Db 2155 ACCGAGCGCCAGTACGCGCTGGGCATCATCCAGCGCCAGCCGACAGAGCGAGCGAG 2214  
QY 2221 CTGCTGACACGATCATCGACGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTG 2280  
Db 2215 CTGCTGACACGATCATCGACGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTG 2274  
QY 2281 CCCGCCCAACAGGGGATCGCGCGCAACGAGCAGATCGCAAGCTGGTGGAGCAAGGGCATC 2340  
Db 2275 CCCGCCCAACAGGGGATCGCGCGCAACGAGCAGATCGCAAGCTGGTGGAGCAAGGGCATC 2334  
QY 2341 CGCAAGGTGCTGTTCTCTGAGCGGCATCGATGGCGGCATCGTGTATCTACCAAGTACATGGAC 2400  
Db 2335 CGCAAGGTGCTGTTCTCTGAGCGGCATCGATGGCGGCATCGTGTATCTACCAAGTACATGGAC 2394  
QY 2401 GACCTGTACGTGGGCGAGCGCGGCGCTTAGGTAGGATCGATTAAAGCTTCCGGGGCTAGCAC 2460  
Db 2395 GACCTGTACGTGGGCGAGCGCGGCGCTTAGGTAGGATCGATTAAAGCTTCCGGGGCTAGCAC 2454  
QY 2461 GGT 2463  
Db 2455 GGT 2457

RESULT 6  
ADCL3265  
ID ADCL3265 standard; DNA; 2457 BP.  
XX AC ADCL3265;  
XX AC ADCL3265;  
DT 18-DEC-2003 (first entry)  
DE DNA of HIV construct p2Pol-opt-ym\_c SEQ ID NO 44.  
XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
KW Rev; HIV type C; Gene therapy; vaccine; immunisation; HIV; ds.  
XX Human immunodeficiency virus.  
XX

PN WO2003004620-A2.  
XX 16-JAN-2003.  
XX 05-JUL-2002; 2002WO-US021420.  
XX 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
XX WPI; 2003-221593/21.  
XX New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.  
XX Disclosure; Fig 41; 301pp; English.  
XX The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intradermally, transdermally, intranasally,  
CC subcutaneously, orally or intravenously. The expression cassette is useful  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC Type C related sequence of the invention.  
XX Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

Query Match 98.6%; Score 2434.6; DB 9; Length 2457;  
Best Local Similarity 99.6%; Pred. No. 9.7e-294;  
Matches 2453; Conservative 0; Mismatches 4; Indels 6; Gaps 1;  
QY 1 GTGAGCGCCACCATGGCGAGGCCATGAGCCAGGCCACAGGCCAACATCTCTGATGAG 60  
Db 1 GTGAGCGCCACCATGGCGAGGCCATGAGCCAGGCCACAGGCCAACATCTCTGATGAG 60  
QY 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCACTGGCGCAAGAGGGC 120  
Db 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCACTGGCGCAAGAGGGC 120  
QY 121 CACATCGCGCGCAACTGCGCGCCCCCGCCAGAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Db 121 CACATCGCGCGCAACTGCGCGCCCCCGCCAGAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
QY 181 GGCCACACAGTGAAGGACTGCGACCGAGCGCCAGGCCAACCTTCTCCGCGAGACCTGGCC 240  
Db 181 GGCCACACAGTGAAGGACTGCGACCGAGCGCCAGGCCAACCTTCTCCGCGAGACCTGGCC 240  
QY 241 TTCCCCCGAGGCAAGGCCCGCGAGTTTCCCGAGGAGAGAACCCGCGCCAGAGCCCGCC 300  
Db 241 TTCCCCCGAGGCAAGGCCCGCGAGTTTCCCGAGGAGAGAACCCGCGCCAGAGCCCGCC 300  
QY 301 AGCCGCGAGCTGAGGTGCGCGGACACACCCCGCAGCGAGCGCGCGCGCGCGCCAG 360  
Db 301 AGCCGCGAGCTGAGGTGCGCGGACACACCCCGCAGCGAGCGCGCGCGCGCGCCAG 360  
QY 361 GGCAACCTGAACCTTCCCCCAGATCACTCTGTGGCAGCGCCCTGTGTGAGCATCAAGTG 420  
Db 361 GGCAACCTGAACCTTCCCCCAGATCACTCTGTGGCAGCGCCCTGTGTGAGCATCAAGTG 420  
QY 421 GGGCGCGAGATCAAGAGGCGCCCTGTGGCCACCGCGCGCGAGACACCGCTGTGGAGGAG 480  
Db 421 GGGCGCGAGATCAAGAGGCGCCCTGTGGCCACCGCGCGCGAGACACCGCTGTGGAGGAG 480



QY 481 ATGAGCTCCCGGCAAGTGGAGCCCAAGATGATCGGGCATCGGGCTTTCATCAAG 540  
Db 481 ATGAGCTCCCGGCAAGTGGAGCCCAAGATGATCGGGCATCGGGCTTTCATCAAG 540  
QY 541 GTGCGCAGTACGACCAAGATCTCTGATCGAGATCTGGGCAAGAAGGCCATCGGACCGGTG 600  
Db 541 GTGCGCAGTACGACCAAGATCTCTGATCGAGATCTGGGCAAGAAGGCCATCGGACCGGTG 600  
QY 601 CTGATCGGCCCCCGGCGTGAACATCATCGGCGCGCAACATGTCGACGAGTGGGCTGC 660  
Db 601 CTGATCGGCCCCCGGCGTGAACATCATCGGCGCGCAACATGTCGACGAGTGGGCTGC 660  
QY 661 ACCCTGAACTTCCCATCATCGACCCCATCGAGACCGTGCCTGGAAGTGAAGCCCGGCGATG 720  
Db 661 ACCCTGAACTTCCCATCATCGACCCCATCGAGACCGTGCCTGGAAGTGAAGCCCGGCGATG 720  
QY 721 GACGCCCCAAGTGAAGAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
Db 721 GACGCCCCAAGTGAAGAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
QY 781 ATCTCGAGGAGATGGAGAGAGGCGCAAGATCACCAAGATCGGCCCCCGAGAACCCCTAC 840  
Db 781 ATCTCGAGGAGATGGAGAGAGGCGCAAGATCACCAAGATCGGCCCCCGAGAACCCCTAC 840  
QY 841 AACACCCCGTGTTCGATCAAG 900  
Db 841 AACACCCCGTGTTCGATCAAG 900  
QY 901 TTCCGCGAGTGAACAGGCGCACCGAGCTTCTGGGAGGTGAGTGGGCGATCCCCCGAC 960  
Db 901 TTCCGCGAGTGAACAGGCGCACCGAGCTTCTGGGAGGTGAGTGGGCGATCCCCCGAC 960  
QY 961 CCGCGCGCTGAAG 1020  
Db 961 CCGCGCGCTGAAG 1020  
QY 1021 AGCTGCGCTGGAG 1080  
Db 1021 AGCTGCGCTGGAG 1080  
QY 1081 AACGAG 1140  
Db 1081 AACGAG 1140  
QY 1141 CCGAGCATCTTCCAG 1200  
Db 1141 CCGAGCATCTTCCAG 1200  
QY 1201 GAGATCGTGTATCTACCAAGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
Db 1201 GAGATCGTGTATCTACCAAGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
QY 1261 CAGACCGCGCCAG 1320  
Db 1261 CAGACCGCGCCAG 1320  
QY 1321 CCGAGCAAG 1380  
Db 1321 CCGAGCAAG 1380  
QY 1381 GACAAAGTGGAG 1440  
Db 1381 GACAAAGTGGAG 1440  
QY 1441 ATCCAG 1500  
Db 1441 ATCCAG 1500  
QY 1501 GCGCAGCTGTGCAAGCTGTGCGGCGCGCCCAAGGCCCTGACCGAGATCTGTCGCCCTGACC 1560  
Db 1501 GCGCAGCTGTGCAAGCTGTGCGGCGCGCCCAAGGCCCTGACCGAGATCTGTCGCCCTGACC 1560

## RESULT 7

ABL39961

ID ABL39961 standard; DNA; 2457 BP.

XX ABL39961;

AC ABL39961;

XX 15-MAY-2002 (first entry)

XX

QY 1561 GAGAGGCCGAGCTGGAGCTGGCCGAGAAACCGCGAGATCTTGGCGAGCCGCTGCAACGCG 1620  
Db 1561 GAGAGGCCGAGCTGGAGCTGGCCGAGAAACCGCGAGATCTTGGCGAGCCGCTGCAACGCG 1620  
QY 1621 GTGTACTTACACCCAGCAGACCTGGTGGCGAGATCCAGAAGCAGAGCCAGACGACGAG 1680  
Db 1621 GTGTACTTACACCCAGCAGACCTGGTGGCGAGATCCAGAAGCAGAGCCAGACGACGAG 1680  
QY 1681 TGGACCTTACAGATCTTACACGAGAGCCCTTCAAGAACCTTGAAGACCCGGCAAGTACGCAAG 1740  
Db 1681 TGGACCTTACAGATCTTACACGAGAGCCCTTCAAGAACCTTGAAGACCCGGCAAGTACGCAAG 1740  
QY 1741 ATGCGCACCGCCACACACAG 1800  
Db 1741 ATGCGCACCGCCACACACAG 1800  
QY 1795 ATGCGCACCGCCACACACAG 1794  
Db 1795 ATGCGCACCGCCACACACAG 1794  
QY 1801 ATGAGAGAGATCTGTGATCTGGGCAAGACCCCAAGTTCGCTTCCGATCCAGAGAGAG 1860  
Db 1795 ATGAGAGAGATCTGTGATCTGGGCAAGACCCCAAGTTCGCTTCCGATCCAGAGAGAG 1854  
QY 1861 ACCTGGAGAGACCTGTGTGAG 1920  
Db 1861 ACCTGGAGAGACCTGTGTGAG 1920  
QY 1915 GTGAACACCCCGCTTGGTGAAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1974  
Db 1915 GTGAACACCCCGCTTGGTGAAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1974  
QY 1981 GCGAGAGACCTTCTACGTTGAG 2040  
Db 1981 GCGAGAGACCTTCTACGTTGAG 2040  
QY 1975 GCGAGAGACCTTCTACGTTGAG 2034  
Db 1975 GCGAGAGACCTTCTACGTTGAG 2034  
QY 2041 TACGTGACCGAG 2100  
Db 2041 TACGTGACCGAG 2100  
QY 2035 TACGTGACCGAG 2094  
Db 2035 TACGTGACCGAG 2094  
QY 2101 ACCGAGCTGCGAG 2160  
Db 2101 ACCGAGCTGCGAG 2160  
QY 2095 ACCGAGCTGCGAG 2154  
Db 2095 ACCGAGCTGCGAG 2154  
QY 2161 ACCGAG 2220  
Db 2161 ACCGAG 2220  
QY 2155 ACCGAG 2214  
Db 2155 ACCGAG 2214  
QY 2221 CTGTGTAAACAGATCATCGAG 2280  
Db 2221 CTGTGTAAACAGATCATCGAG 2280  
QY 2215 CTGTGTAAACAGATCATCGAG 2274  
Db 2215 CTGTGTAAACAGATCATCGAG 2274  
QY 2281 CCGGCGCCAG 2340  
Db 2281 CCGGCGCCAG 2340  
QY 2275 CCGGCGCCAG 2334  
Db 2275 CCGGCGCCAG 2334  
QY 2341 CGCAAGGTGCTGTTCTTCTGAG 2400  
Db 2341 CGCAAGGTGCTGTTCTTCTGAG 2400  
QY 2335 CGCAAGGTGCTGTTCTTCTGAG 2394  
Db 2335 CGCAAGGTGCTGTTCTTCTGAG 2394  
QY 2401 GACCTGTGCTGGGAG 2460  
Db 2401 GACCTGTGCTGGGAG 2460  
QY 2395 GACCTGTGCTGGGAG 2454  
Db 2395 GACCTGTGCTGGGAG 2454  
QY 2461 GGT 2463  
Db 2455 GGT 2457



Db	1429	ATCCAGAGCTGTGGGCAAGCTGAACCTGGCCAGCCAGATCTACCCCGGCATCAAGGTG	1488
QY	1501	CGCCAGCTGTGAAGCTGTGGCGCGCCAGAGGCGCTGACCGACATCGTGCCCTGACC	1560
Db	1489	CGCCAGCTGTGAAGCTGTGGCGCGCCAGAGGCGCTGACCGACATCGTGCCCTGACC	1548
QY	1561	GAGGAGCCGAGCTGAGCTGGCCGAGAACCCGCGAGATCCTCGCGAGCCCGTGACCGC	1620
Db	1549	GAGGAGCCGAGCTGAGCTGGCCGAGAACCCGCGAGATCCTCGCGAGCCCGTGACCGC	1608
QY	1621	GTGTACTACACCCGAGCAGACCTGGTGGCGGAGATCCAGAACGAGGCGCCACACAG	1680
Db	1609	GTGTACTACACCCGAGCAGACCTGGTGGCGGAGATCCAGAACGAGGCGCCACACAG	1668
QY	1681	TGGACCTACAGATCTACACGAGGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG	1740
Db	1669	TGGACCTACAGATCTACACGAGGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG	1728
QY	1741	ATCGCAGCCGCCACACCAACGACGCTGAAGAGCTGACCGAGCGCGTGAGAGATCGCC	1800
Db	1729	ATCGCAGCCGCCACACCAACGACGCTGAAGAGCTGACCGAGCGCGTGAGAGATCGCC	1788
QY	1801	ATGAGAGATCTGTATCTGGGCGAGACCCCGCAAGTTCGGCTGCCCATCCAGAGAGAG	1860
Db	1789	ATGAGAGATCTGTATCTGGGCGAGACCCCGCAAGTTCGGCTGCCCATCCAGAGAGAG	1848
QY	1861	ACCTGGAGACCTGTGTGACCGACTACTGGCAGGCCACTGTGATCCCGAGTGGAGTTC	1920
Db	1849	ACCTGGAGACCTGTGTGACCGACTACTGGCAGGCCACTGTGATCCCGAGTGGAGTTC	1908
QY	1921	GTGAACACCCCGCTGTGTGAGCTGTGTACCGCTGAGAGGAGCCCATCATCGCG	1980
Db	1909	GTGAACACCCCGCTGTGTGAGCTGTGTACCGCTGAGAGGAGCCCATCATCGCG	1968
QY	1981	GCCGAGACCTTCTACCTGTGACCGCGCCGCAACCCGAGACCAAGATCGGCAAGCCGCG	2040
Db	1969	GCCGAGACCTTCTACCTGTGACCGCGCCGCAACCCGAGACCAAGATCGGCAAGCCGCG	2028
QY	2041	TACGTGACCGACCGGCGCGGAGAGATCGTGAGCTGACCGAGACCAACCAAGAG	2100
Db	2029	TACGTGACCGACCGGCGCGGAGAGATCGTGAGCTGACCGAGACCAACCAAGAG	2088
QY	2101	ACCGAGCTGACGCGCCATCCAGCTGGCCCTGCAGGACAGCGGCGAGGTGAACATCGTG	2160
Db	2089	ACCGAGCTGACGCGCCATCCAGCTGGCCCTGCAGGACAGCGGCGAGGTGAACATCGTG	2148
QY	2161	ACCGACGCGAGTACGCTTGGGCGATCATCCAGGCCCGCCGACAGAGCGAGCGAG	2220
Db	2149	ACCGACGCGAGTACGCTTGGGCGATCATCCAGGCCCGCCGACAGAGCGAGCGAG	2208
QY	2221	CTGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTG	2280
Db	2209	CTGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTG	2268
QY	2281	CCCGCCACAGGCGCATCGCGCGCAACGAGCAGATCGAACGCTGTGAGCGAGGCGATC	2340
Db	2269	CCCGCCACAGGCGCATCGCGCGCAACGAGCAGATCGAACGCTGTGAGCGAGGCGATC	2328
QY	2341	CGCAAGTGCTGTCTCTGAGCGGCATCGATGGCGGATCGTATCTACCAATACATGGAC	2400
Db	2329	CGCAAGTGCTGTCTCTGAGCGGCATCGATGGCGGATCGTATCTACCAATACATGGAC	2388
QY	2401	GACCTGTACCTGGGCGCGCGCTTGTAGATCGATTAAGCTTCCCGGGGCTAGCAC	2460
Db	2389	GACCTGTACCTGGGCGCGCGCTTGTAGATCGATTAAGCTTCCCGGGGCTAGCAC	2448
QY	2461	GGTGAATTC	2469
Db	2449	GGTGAATTC	2457

ID XX ACA03546 standard; DNA; 2445 BP.  
AC ACA03546;  
XX  
DT 22-MAY-2003 (first entry)  
XX  
DE Synthetic DNA encoding immunogenic HIV peptide #29.  
XX  
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
KW gene therapy; packaging cell line; humoral immune response;  
KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
XX  
OS Synthetic.  
XX  
FN W02003004657-A1.  
XX  
PD 16-JAN-2003.  
XX  
PP 05-JUL-2002; 2002WO-US021421.  
XX  
PR 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349728P.  
PR 16-JAN-2002; 2002US-0349793P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
FI Zur Megede J, Barnett SW, Lian Y;  
XX  
DR WPI; 2003-221602/21.  
XX  
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
PT polypeptides, useful as immunogenic compositions or vaccines for  
PT generating humoral or cellular immune responses against HIV in a subject,  
PT especially humans.  
XX  
PS Example 1; Fig 34; 262pp; English.  
XX  
CC The invention describes a synthetic polynucleotide encoding 2 or more  
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
CC derived from different HIV subtypes. The polynucleotide is useful for  
CC immunisation, generation of packaging cell lines, or production of HIV  
CC polypeptides. The polynucleotide and its encoded proteins are useful as  
CC immunogenic compositions or vaccines for generating humoral or cellular  
CC immune responses against HIV in a subject, or for inducing neutralising  
CC antibodies against HIV. The gene delivery vector comprising the  
CC polynucleotide is also useful for DNA immunisation of, or for generating  
CC an immune response (e.g. a humoral or cellular immune response), in, a  
CC subject such as a mammal, particularly a human. This sequence encodes a  
CC human immunodeficiency virus immunogenic peptide  
XX  
SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match	97.3%;	Score	2401.8;	DB	7;	Length	2445;
Best Local Similarity	99.2%;	Pred. No.	1.1e-289;				
Matches	2438;	Conservative	0;	Mismatches	7;	Indels	12;
Gaps	2						
QY	7	GCCACCATGGCCGAGGCGCATGAGCCAGGCCACAGCGCCCAACATCTCTGATGCGAGCGCAGC	66				
Db	1	GCCACCATGGCCGAGGCGCATGAGCCAGGCCACAGCGCCCAACATCTCTGATGCGAGCGCAGC	60				
QY	67	AACCTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTCAACTGCGCAAGGAGGGCCACATC	126				
Db	61	AACCTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGGCCACATC	120				
QY	127	GCCCGCAACTGCGCGCGCCCCCGCAAGAGGGCTGTCTGGAAGTGTGCGCAAGGAGGGCCAC	186				
Db	121	GCCCGCAACTGCGCGCGCCCCCGCAAGAGGGCTGTCTGGAAGTGTGCGCAAGGAGGGCCAC	180				
QY	187	CAGATGAAGGACTGCAAGGAGGGCCAGGCCAACTTCTTCCGCGAGGACTGTGCGCTTCCCG	246				
Db	181	CAGATGAAGGACTGCAAGGAGGGCCAGGCCAACTTCTTCCGCGAGGACTGTGCGCTTCCCG	240				















Qy	74	AGGGCCCAAGCGCATCATCAAGTGTCTTAACTTCGGCAAGAGAGGGCCA	CATCGCCGCGCA	133
Db	1547	AGGGCCCAAGCGCATCATCAAGTGTCTTAACTTCGGCAAGAGAGGGCCA	CATCGCCGCGCA	1606
Qy	134	ACTCGCGGCCCCCGCAAGAAGGGCTGTGGAAAGTGGCGCAAGAGAGGGCCA	CAAGATGA	193
Db	1607	ACTCGCGGCCCCCGCAAGAGGGCTGTGGAAAGTGGCGCAAGAGAGGGCCA	CAAGATGA	1666
Qy	194	AGGACTGACCGAGCGCCAGCCAACTTCTTCGCGAGGACCTGGGCTT	CCCCCAGGGCA	253
Db	1667	AGGACTGACCGAGCGCCAGCCAACTTCTTCGCGAGGACCTGGGCTT	CCCCCAGGGCA	1726
Qy	254	AGGCCCGGAGTTCCCCAGCGAGCAGAACCGCGCCAAACAGCCCAACAGCGCGGAGGTGC		313
Db	1727	AGGCCCGGAGTTCCCCAGCGAGCAGAACCGCGCCAAACAGCCCAACAGCGCGGAGGTGC		1786
Qy	314	AGGTGCGCGCGCAAAACCCCGCAGCGAGGCGCGCGCGCGAGCGCCAGGGCAACCCCTGAAC		373
Db	1787	AGGTGCGCGCGCAAAACCCCGCAGCGAGGCGCGCGCGCGAGCGCCAGGGCAACCCCTGAAC		1846
Qy	374	TCCCCCAGATCACCTGTGCGAGCGCCCTGTGTGAGCATCAAGTGGGCGGCCAGATCA		433
Db	1847	TCCCCCAGATCACCTGTGCGAGCGCCCTGTGTGAGCATCAAGTGGGCGGCCAGATCA		1906
Qy	434	AGGAGGCCCTGCTGGACACCGCGCCCGACACACCGTGTCTGGAGGAGATCAGAGCTTGCCCG		493
Db	1907	AGGAGGCCCTGCTGGACACCGCGCCCGACACACCGTGTCTGGAGGAGATCAGAGCTTGCCCG		1966
Qy	494	GCAAGTGGAGGCCAAGATGATTCGGCGGCATTCGGCGGCTTCATCAAGTGGCGCCAGTACG		553
Db	1967	GCAAGTGGAGGCCAAGATGATTCGGCGGCATTCGGCGGCTTCATCAAGTGGCGCCAGTACG		2026
Qy	554	ACCAGATCTCTGATCGAGATCTCGCGCAAGAGGCCATCGGCACCGTCTCTCATCGGCCCA		613
Db	2027	ACCAGATCTCTGATCGAGATCTCGCGCAAGAGGCCATCGGCACCGTCTCTCATCGGCCCA		2086
Qy	614	CCCCCGTGAAACATCATCGCGCGCAACATGTGACCCAGCTGGGCTGACCCCTGAACCTTC		673
Db	2087	CCCCCGTGAAACATCATCGCGCGCAACATGTGACCCAGCTGGGCTGACCCCTGAACCTTC		2146
Qy	674	CCATCAGCCCATCGAGACCGTGCCTGTGAAGTGAAGCCCGGCATGGACGGGCCCAAGG		733
Db	2147	CCATCAGCCCATCGAGACCGTGCCTGTGAAGTGAAGCCCGGCATGGACGGGCCCAAGG		2206
Qy	734	TGAAGCAGTGGCCCTTACCGAGGAGAGATCAAGGCCCTGACCGCCATCTCGAGGAGA		793
Db	2207	TGAAGCAGTGGCCCTTACCGAGGAGAGATCAAGGCCCTGACCGCCATCTCGAGGAGA		2266
Qy	794	TGGAGAGGAGGGCAAGATCAACAGATTCGGCCCGGAGAACCCCTCAACAACCCCGGTGT		853
Db	2267	TGGAGAGGAGGGCAAGATCAACAGATTCGGCCCGGAGAACCCCTCAACAACCCCGGTGT		2326
Qy	854	TCGGCATCAAGAAGAAGCAGCAGCAACAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGA		913
Db	2327	TCGGCATCAAGAAGAAGCAGCAGCAACAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGA		2386
Qy	914	ACAAGCGCACCCAGGACTTTCGGAGGTGCAGTGGGCATCCCCCACC	CCCGCGGCTGA	973
Db	2387	ACAAGCGCACCCAGGACTTTCGGAGGTGCAGTGGGCATCCCCCACC	CCCGCGGCTGA	2446
Qy	974	AGAGNAGNAGAGCGTCAACCGTGTGACGTTGGCGAGCGCTACTTCAGCGTGCCTCTGG		1033
Db	2447	AGAGNAGNAGAGCGTCAACCGTGTGACGTTGGCGAGCGCTACTTCAGCGTGCCTCTGG		2506
Qy	1034	ACGAGGACTTCCGCAAGTACACCGCCTTACCACTCCCCAGCATCAACAAGAGACCCCGG		1093
Db	2507	ACGAGGACTTCCGCAAGTACACCGCCTTACCACTCCCCAGCATCAACAAGAGACCCCGG		2566
Qy	1094	GCATCCCGTACCAATACAAAGTGTGCTGCCCGAGGCTGGAGGGGAGGCCCAGCATCTTCC		1153
Db	2567	GCATCCCGTACCAATACAAAGTGTGCTGCCCGAGGCTGGAGGGGAGGCCCAGCATCTTCC		2626

Qy	1154	AGAGCAGCATGACCAAGATCCTTGAGGCCCTTCGCGCCCGCAACCCCGAGATCGTGATCT	1211
Db	2627	AGAGCAGCATGACCAAGATCCTTGAGGCCCTTCGCGCCCGCAACCCCGAGATCGTGATCT	2686
Qy	1214	ACCAGTACATGAGCAGCCTCTAGCTGGGCAAGCACTTGAGATCGGCCAGCACCGCGCCA	1273
Db	2687	ACCA-----GGCCCCCTGTACTGGGCAAGCACTTGAGATCGGCCAGCACCGCGCCA	2740
Qy	1274	AGATCGAGAGCTGCGCAAGCACTGTGTGGCTGGGGCTTTCACACCCCCGACAGAAAGC	1333
Db	2741	AGATCGAGAGCTGCGCAAGCACTGTGTGGCTGGGGCTTTCACACCCCCGACAGAAAGC	2800
Qy	1334	ACCAGAAGGAGCCCCCTTCTCTGTGATGGCTACGAGCTGCACCCCGACAAGTGGAACCG	1393
Db	2801	ACNAGAAGGAGCCCCCTTCTCTGCCAT-----CGAGCTGCACCCCGACAAGTGGAACCG	2854
Qy	1394	TGCAGCCCATCGAGTGTCCCGAAGAGAGCTGGACCTGTGAACGACATCCAGAAAGCTGG	1453
Db	2855	TGCAGCCCATCGAGTGTCCCGAAGAGAGCTGGACCTGTGAACGACATCCAGAAAGCTGG	2914
Qy	1454	TGGCAGACTGAACTGGGCGAGCCAGATCTACCCCGGCATCAAGTGTGGCCAGCTGTGCA	1513
Db	2915	TGGCAGACTGAACTGGGCGAGCCAGATCTACCCCGGCATCAAGTGTGGCCAGCTGTGCA	2974
Qy	1514	AGCTGTCTCGCGCGCCCAAGGCCCTTGACGACATCTGTGCCCTTACCGAGAGGCCGAGC	1573
Db	2975	AGCTGTCTCGCGCGCCCAAGGCCCTTGACGACATCTGTGCCCTTACCGAGAGGCCGAGC	3034
Qy	1574	TGGAGCTGGCGAGAACCCGCGAGATCCTGTGCGAGCCCGTGCACGGCGTGTACTACGACC	1633
Db	3035	TGGAGCTGGCGAGAACCCGCGAGATCCTGTGCGAGCCCGTGCACGGCGTGTACTACGACC	3094
Qy	1634	CCAGCAAGGACCTGGTGGCCGAGATCCAGAAAGCAGGGCCACGACCACTGAGACCTACCAGA	1693
Db	3095	CCAGCAAGGACCTGGTGGCCGAGATCCAGAAAGCAGGGCCACGACCACTGAGACCTACCAGA	3154
Qy	1694	TCTACAGGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACGCCAAGATCGCGACCGCCC	1753
Db	3155	TCTACAGGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACGCCAAGATCGCGACCGCCC	3214
Qy	1754	ACACCAAGCAGTGAAGCAGCTGACCGAGGCCGTGCAGAAAGATCGCCATCGAGAGCATCG	1813
Db	3215	ACACCAAGCAGTGAAGCAGCTGACCGAGGCCGTGCAGAAAGATCGCCATCGAGAGCATCG	3274
Qy	1814	TGATCTGGGGCAAGACCCCAAGTTCCGCTTGCCCATCCAGAAAGGACCTGGGAGACCT	1873
Db	3275	TGATCTGGGGCAAGACCCCAAGTTCCGCTTGCCCATCCAGAAAGGACCTGGGAGACCT	3334
Qy	1874	GGTGACCGCATCTGCGACGGCCACTGGATTCGCCAGTGGGAGTTCTGTAAACACCCCCC	1933
Db	3335	GGTGACCGCATCTGCGACGGCCACTGGATTCGCCAGTGGGAGTTCTGTAAACACCCCCC	3394
Qy	1934	CCCTGTGTAACTGTGGTACACAGCTGGAGAAGGAGCCCATCATCGCGCCCGAGACCTTCT	1993
Db	3395	CCCTGTGTAACTGTGGTACACAGCTGGAGAAGGAGCCCATCATCGCGCCCGAGACCTTCT	3454
Qy	1994	ACGTGAGCGGCGGCCCAACCGCGAGACCAAGATCGCAAGGCCCGGTGTAAGTGCAGC	2053
Db	3455	ACGTGAGCGGCGGCCCAACCGCGAGACCAAGATCGCAAGGCCCGGTGTAAGTGCAGC	3514
Qy	2054	GGGCGCGGAGAGATCGTGAGCTGACCGAGACCACCAACCAAGAACCGAGCTGCAGG	2113
Db	3515	GGGCGCGGAGAGATCGTGAGCTGACCGAGACCACCAACCAAGAACCGAGCTGCAGG	3574
Qy	2114	CCATCCAGCTGGCCCTTCAGGACAGCGGCGAGCGAGGTGAACATCTGTGACCAAGCCAGT	2173
Db	3575	CCATCCAGCTGGCCCTTCAGGACAGCGGCGAGCGAGGTGAACATCTGTGACCAAGCCAGT	3634
Qy	2174	ACGCCCTGGGCATCATCCAGGCCACCGCCACAAGACGAGAGCGAGCTGGTAGACAGA	2233
Db	3635	ACGCCCTGGGCATCATCCAGGCCACCGCCACAAGACGAGAGCGAGCTGGTAGACAGA	3694
Qy	2234	TCATCGAGCAGCTGATCAAGAAAGGAAAGGTGTACTCTGAGCTGGGTGCCCGCCCAAGG	2293

3695 TCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGTGGTGGCCGCCACAAAG 3754  
 2294 GCATCGCGCGGACAGGACGATCGAAGCTGTGTGAGCAAGGCGATCCGCAAGTCTGT 2353  
 3755 GCATCGCGCGGACAGGACGATCGAAGCTGTGTGAGCAAGGCGATCCGCAAGTCTGT 3814  
 2354 TCCTGGACGGGATCGATGGCGGATCGTGTATCTACAGTACATGGACGACCTGTACGTGG 2413  
 3815 TCCTGGACGGGATCGATGGCGGATCGTGTATCTACAGTACATGGACGACCTGTACGTGG 3874  
 2414 GCAGCGCGGCGCTTAGATCGATTAAAGCTTCCGGGGGTACACGGT 2463  
 3875 GCAGCGCGGCGCTTAGATCGATTAAAGCTTCCGGGGGTACACGGT 3924

RESULT 13

ACA03591

ID ACA03591 standard; DNA; 5184 BP.

AC AC

XX ACA03591;

BT 22-MAY-2003 (first entry)

XX Synthetic DNA encoding immunogenic HIV peptide #74.

DE Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;

KW gene therapy; packaging cell line; humoral immune response;

KW cellular immune response; gene delivery vector; DNA immunisation; ds.

XX OS Synthetic.

XX WO2003004657-A1.

PN 16-JAN-2003.

XX 05-JUL-2002; 2002WO-US021421.

XX 05-JUL-2001; 2001US-0303192P.

PR 31-AUG-2001; 2001US-0316860P.

PR 16-JAN-2002; 2002US-0349728P.

PR 16-JAN-2002; 2002US-0349793P.

PR 16-JAN-2002; 2002US-0349871P.

XX (CHIR ) CHIRON CORP.

XX Zur Megede J, Barnett SW, Lian Y;

XX WPI; 2003-221602/21.

XX New synthetic polynucleotides encoding antigenic HIV type B and/or type C

PT polypeptides, useful as immunogenic compositions or vaccines for

PT generating humoral or cellular immune responses against HIV in a subject,

PT especially humans.

XX Example 1; Fig 79; 262pp; English.

XX The invention describes a synthetic polynucleotide encoding 2 or more

CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are

CC derived from different HIV subtypes. The polynucleotide is useful for

CC immunisation, generation of packaging cell lines, or production of HIV

CC polypeptides. The polynucleotide and its encoded proteins are useful as

CC immunogenic compositions or vaccines for generating humoral or cellular

CC immune responses against HIV in a subject, or for inducing neutralising

CC antibodies against HIV. The gene delivery vector comprising the

CC polynucleotide is also useful for DNA immunisation of, or for generating

CC an immune response (e.g. a humoral or cellular immune response) in, a

CC subject such as a mammal, particularly a human. This sequence encodes a

CC human immunodeficiency virus immunogenic peptide

XX Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;

XX Query Match 96.9%; Score 2393.2; DB 7; Length 5184;

Best Local Similarity 99.2%; Pred. No. 1.2e-288;  
 Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;  
 QY 14 TGGCGGAGGCGCATCAGCGAGGCGCCACAGCGCCAAACATCTCTGATGCGAGCGCAGCAACTTCA 73  
 DB 2741 TGGCGGAGGCGCATCAGCGAGGCGCCACAGCGCCAAACATCTCTGATGCGAGCGCAGCAACTTCA 2800  
 QY 74 AGGSCCCCAAGCGCATCATCAAGTGTCTTCACTGCGGCAAGGAGGCGCAATCGCCGCGCA 133  
 DB 2801 AGGSCCCCAAGCGCATCATCAAGTGTCTTCACTGCGGCAAGGAGGCGCAATCGCCGCGCA 2860  
 QY 134 ACTGCGCGCGCCCGGCAAGAGGCTGCTGGAGTGGCGCAAGAGGAGGCGCAACAGATGA 193  
 DB 2861 ACTGCGCGCGCCCGGCAAGAGGCTGCTGGAGTGGCGCAAGAGGAGGCGCAACAGATGA 2920  
 QY 194 AGGACTGCACCGAGCGCGCAGCCAACTTTCTTCGCGAGGACCTGGCTTTCCCCAGGGCA 253  
 DB 2921 AGGACTGCACCGAGCGCGCAGCCAACTTTCTTCGCGAGGACCTGGCTTTCCCCAGGGCA 2980  
 QY 254 AGGCGCGCGAGTTCCCGAGCGAGCAGAACCGCGCCACACAGCCCGCCAGCGCGGAGCTGC 313  
 DB 2981 AGGCGCGCGAGTTCCCGAGCGAGCAGAACCGCGCCACACAGCCCGCCAGCGCGGAGCTGC 3040  
 QY 314 AGGTGCGCGCGCACAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
 DB 3041 AGGTGCGCGCGCACAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3100  
 QY 374 TCCCGCAGATCACCCTGTGGCAGCGCGCCCTGTGTAGCATCAAGTGGCGGCGCAGATCA 433  
 DB 3101 TCCCGCAGATCACCCTGTGGCAGCGCGCCCTGTGTAGCATCAAGTGGCGGCGCAGATCA 3160  
 QY 434 AGGAGGCGCTGTGGACACCG 493  
 DB 3161 AGGAGGCGCTGTGGACACCG 3220  
 QY 494 GCAAGTGAAGCCCAAGATGATCGGCGCGCATCGGCGCGCATCGGCGCGCATCGGCGCGCGCGCG 553  
 DB 3221 GCAAGTGAAGCCCAAGATGATCGGCGCGCATCGGCGCGCATCGGCGCGCATCGGCGCGCGCG 3280  
 QY 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCGCGCATCGGCGCGCGCGCGCG 613  
 DB 3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCGCGCATCGGCGCGCGCGCGCG 3340  
 QY 614 CCCCCGTGAACATCATCGGCGCGCAACATGCTGAGCCAGCTGGGTGCGACCCCTGAACTTCC 673  
 DB 3341 CCCCCGTGAACATCATCGGCGCGCAACATGCTGAGCCAGCTGGGTGCGACCCCTGAACTTCC 3400  
 QY 674 CCATCAGCCCGCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATCGACGGCGCCCAAGG 733  
 DB 3401 CCATCAGCCCGCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATCGACGGCGCCCAAGG 3460  
 QY 734 TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCGCATCTGCGAGGAGA 793  
 DB 3461 TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCGCATCTGCGAGGAGA 3520  
 QY 794 TGGAGAAGGAGGCGCAAGATCACCAAGATCGGCGCGCGAGAGCCCTTACAACACCCCGCTGT 853  
 DB 3521 TGGAGAAGGAGGCGCAAGATCACCAAGATCGGCGCGCGAGAGCCCTTACAACACCCCGCTGT 3580  
 QY 854 TCGCCATCAAGAGAGGAGGAGCAGCAACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGTGA 913  
 DB 3581 TCGCCATCAAGAGAGGAGGAGCAGCACAAGTGGCGCAAGCTGGTGGACTTCGCGAGTGA 3640  
 QY 914 ACAAGGCG 973  
 DB 3641 ACAAGGCG 3700  
 QY 974 AGAAGAAGAGAGCGCTGACCCGCTGTGGAGCTGGCGGCGAGCCCTTACTTCAAGCTGCGCCCTGG 1033  
 DB 3701 AGAAGAAGAGAGCGCTGACCCGCTGTGGAGCTGGCGGCGAGCGCCCTTACTTCAAGCTGCGCCCTGG 3760  
 QY 1034 ACAGAGCTTCGCGAAGTACACCGCTTCAACATCCCGAGCATCAAGAGAGGAGCGCGCGCG 1093

Db 3761 ACGAGGACTTCGGCAAGTACACCGCTTCACCATCCCGAGCATCAACAGAGACCCCG 3820  
QY 1094 GCATCCGCTACCAAGTACAACTGTGCTGCCAGGCTGGAAGGCGAGCCCGAGCATCTTCC 1153  
Db 3821 GCATCCGCTACCAAGTACAACTGTGCTGCCAGGCTGGAAGGCGAGCCCGAGCATCTTCC 3880  
QY 1154 AGAGCAGCATCAACAAAGTCTGTGAGCCCTTCCGGCCCGCAACCCCGAGATCGTGATCT 1213  
Db 3881 AGAGCAGCATCAACAAAGTCTGTGAGCCCTTCCGGCCCGCAACCCCGAGATCGTGATCT 3940  
QY 1214 ACCAGTACATGAGCAGCTGTACGTGGGCGAGCAGCTGTGAGATCGGCGAGCAGCCGCCA 1273  
Db 3941 ACCA-----GGCCCCCTGTACGTGGGCGAGCAGCTGTGAGATCGGCGAGCAGCCGCCA 3994  
QY 1274 AGATCGAGGAGCTGGCAAGCAGCTGTGCTGGGCTTACCACCCCGCAAGAGC 1333  
Db 3995 AGATCGAGGAGCTGGCAAGCAGCTGTGCTGGGCTTACCACCCCGCAAGAGC 4054  
QY 1334 ACCAAGAGAGCCCGCTTCTGTGGATGGCTACGAGCTCACCCCGCAAGTGGACCG 1393  
Db 4055 ACCAAGAGAGCCCGCTTCTGTGGCAT-----CGAGCTGCACCCCGCAAGTGGACCG 4108  
QY 1394 TGCAGCCCATGAGTGCCTCCAGAGAGAGAGCTGACCGTGAAAGCATTCAGAGCTGG 1453  
Db 4109 TGCAGCCCATGAGTGCCTCCAGAGAGAGAGCTGACCGTGAAAGCATTCAGAGCTGG 4168  
QY 1454 TGGGCAAGCTGAATGGGCGCAGCAGATCTACCCCGGCATCAAGGTGGCGAGCTGTGCA 1513  
Db 4169 TGGGCAAGCTGAATGGGCGCAGCAGATCTACCCCGGCATCAAGGTGGCGAGCTGTGCA 4228  
QY 1514 AGCTGCTGGGCGGCGCAAGCCCTGACCGACATCTGTCCCTCAACGAGAGCCGAGC 1573  
Db 4229 AGCTGCTGGGCGGCGCAAGCCCTGACCGACATCTGTCCCTCAACGAGAGCCGAGC 4288  
QY 1574 TGGAGCTGGCGAGAAACCGCAGATCTCTGGCGAGCCCGTGCACGGGTGTACTACGACC 1633  
Db 4289 TGGAGCTGGCGAGAAACCGCAGATCTCTGGCGAGCCCGTGCACGGGTGTACTACGACC 4348  
QY 1634 CCAGCAAGAGCTGTGTGCGGAGATTCAGAGAGCGAGGCCACGACAGTGGACCTTACCAGA 1693  
Db 4349 CCAGCAAGAGCTGTGTGCGGAGATTCAGAGAGCGAGGCCACGACAGTGGACCTTACCAGA 4408  
QY 1694 TCTACGAGAGCCCTTCAAGAACCTTGAAGCCGCAAGTACGCAAGATGGCAGCCGCC 1753  
Db 4409 TCTACGAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCAAGATGGCAGCCGCC 4468  
QY 1754 ACACCAAGCAGTGAAGCAGCTGACCGGAGCGCTGCAGAGATCGCCATGGAGAGCATCG 1813  
Db 4469 ACACCAAGCAGTGAAGCAGCTGACCGGAGCGCTGCAGAGATCGCCATGGAGAGCATCG 4528  
QY 1814 TGATCTGGGCGAGACCCCGAAGTTCCGCTTCCCATCCAGAGAGACCTGGAGACCT 1873  
Db 4529 TGATCTGGGCGAGACCCCGAAGTTCCGCTTCCCATCCAGAGAGACCTGGAGACCT 4588  
QY 1874 GGTGACCGAGTACTTGGCAGGCGCACTGGATCCCGAGTGGAGTTCTGTGAACACCCGCC 1933  
Db 4589 GGTGACCGAGTACTTGGCAGGCGCACTGGATCCCGAGTGGAGTTCTGTGAACACCCGCC 4648  
QY 1934 CCCTGCTGAAGTGTGGTACAGCTGAGAGAGAGCCCATCATCGGCGCCGAGACCTTCT 1993  
Db 4649 CCCTGCTGAAGTGTGGTACAGCTGAGAGAGAGCCCATCATCGGCGCCGAGACCTTCT 4708  
QY 1994 AGTGGAGCGGCGCCCAACCGCAGACCAAGATCGGACGCGCGTACTGTGACCGACC 2053  
Db 4709 AGTGGAGCGGCGCCCAACCGCAGACCAAGATCGGACGCGCGTACTGTGACCGACC 4768  
QY 2054 GGGCGCGCAGAGATCTGTGACCTTGAACCGAGACCAACCAAGAGACCGAGTGGAGG 2113  
Db 4769 GGGCGCGCAGAGATCTGTGACCTTGAACCGAGACCAACCAAGAGACCGAGTGGAGG 4828  
QY 2114 CCATCCAGCTGCTGCGAGCAGCAGCGAGGTGAACATGCTGACGACAGCCAGT 2173  
Db 4829 CCATCCAGCTGCTGCGAGCAGCAGCGAGGTGAACATGCTGACCGAGCCAGT 4888

## RESULT 14

ADCI3279  
ID ID ADCI3279 standard; DNA; 5184 BP.  
XX AC ADCI3279;  
XX AC AC  
XX AC AC  
DT 18-DEC-2003 (first entry)  
XX DE DNA of HIV construct TaRevNefgagCpolIna\_C SEQ ID NO 58.  
XX XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
XX XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
XX OS Human immunodeficiency virus.  
XX XX WO2003004620-A2.  
XX PD 16-JAN-2003.  
XX PF 05-JUL-2002; 2002WO-US021420.  
XX PR 05-JUL-2001; 2001US-0303192P.  
XX PR 31-AUG-2001; 2001US-0318660P.  
XX PR 16-JAN-2002; 2002US-0349871P.  
XX XX (CHIR ) CHIRON CORP.  
XX PA (UYST-) UNIV STELLENBOSCH.  
XX PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
XX DR WPI; 2003-221593/21.  
XX XX New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
PT Prot., or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.  
XX XX Disclosure; Fig 55; 301pp; English.  
XX XX The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intramuscularly, intravenously,  
CC subcutaneously, intradermally, transdermally, intravaginally,  
CC intrarectally, orally or intravenously. The expression cassette is useful  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC Type C related sequence of the invention.  
XX XX

50	Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;	
	Query Match 96.9%; Score 2393.2; DB 9; Length 5184;	
	Best Local Similarity 99.2%; Pred. No. 1.2e-288;	
	Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;	
Qy	14 TGGCCGAGGCGATGAGCCAGCCACACAGCGCCAAACATCTGTGATGAGCGCAGCAACTTCA 73	
Db	2741 TCGCCGAGGCGATGAGCCAGCCACACAGCGCCAAACATCTGTGATGAGCGCAGCAACTTCA 2800	
Qy	74 AGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGCGCCACATCGCCCGCA 133	
Db	2801 AGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGCGCCACATCGCCCGCA 2860	
Qy	134 ACTGCGCGCGCCCGCCGCAAGAGGGCTGTGGAAGTGCAGCAAGGAGGCGCCACAGATGA 193	
Db	2861 ACTGCGCGCGCCCGCCGCAAGAGGGCTGTGGAAGTGCAGCAAGGAGGCGCCACAGATGA 2920	
Qy	194 AGGACTGCACGAGCGCGCCGCAAGTCTTTCGCGAGGACCTTGGCTTCCCGCCAGCGCA 253	
Db	2921 AGGACTGCACGAGCGCGCCGCAAGTCTTTCGCGAGGACCTTGGCTTCCCGCCAGCGCA 2980	
Qy	254 AGGCCCCGAGTTCCTCCAGCGAGAGAAACCGCGCCAAACAGCCCAACAGCGCGAGCTGC 313	
Db	2981 AGGCCCCGAGTTCCTCCAGCGAGAGAAACCGCGCCAAACAGCCCAACAGCGCGAGCTGC 3040	
Qy	314 AGGTGCGCGCGCAACACCCCGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373	
Db	3041 AGGTGCGCGCGCAACACCCCGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3100	
Qy	374 TCCCGCCAGATCACCTGTGCGAGCGCCCTGTGTGAGCATCAAGTGGGCGCGCAGATCA 433	
Db	3101 TCCCGCCAGATCACCTGTGCGAGCGCCCTGTGTGAGCATCAAGTGGGCGCGCAGATCA 3160	
Qy	434 AGGAGGCGCTGTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493	
Db	3161 AGGAGGCGCTGTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3220	
Qy	494 GCAAGTGGAGCCCAAGATGATGCGCGCATGCGCGCATGCGCGCATGCGCGCATGCGCG 553	
Db	3221 GCAAGTGGAGCCCAAGATGATGCGCGCATGCGCGCATGCGCGCATGCGCGCATGCGCG 3280	
Qy	554 ACCAGTCTCGATCGAGATCTGCGGCAAGAGGCGCATGCGCGCATGCGCGCATGCGCGCG 613	
Db	3281 ACCAGTCTCGATCGAGATCTGCGGCAAGAGGCGCATGCGCGCATGCGCGCATGCGCGCG 3340	
Qy	614 CCCCCTGTGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGGCTGACCCCTGAACTTC 673	
Db	3341 CCCCCTGTGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGGCTGACCCCTGAACTTC 3400	
Qy	674 CCATCAGCCCATCGAGACCGTGCCTGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 733	
Db	3401 CCATCAGCCCATCGAGACCGTGCCTGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 3460	
Qy	734 TGAAGCAGTGGCCCTGACCCGAGGAGAGATCAAGGCGCTGACCGCCATCTGCGAGGAGA 793	
Db	3461 TGAAGCAGTGGCCCTGACCCGAGGAGAGATCAAGGCGCTGACCGCCATCTGCGAGGAGA 3520	
Qy	794 TGGAGAGGAGGCGCAAGATCAACAGATCGGCGCCCGAGAACCCCTTCAACACCCCGTGT 853	
Db	3521 TGGAGAGGAGGCGCAAGATCAACAGATCGGCGCCCGAGAACCCCTTCAACACCCCGTGT 3580	
Qy	854 TCGCCATCAAG 913	
Db	3581 TCGCCATCAAG 3640	
Qy	914 ACAAGCGCACCCAGAGATCTTGGAGAGTGCAGCTGGGCGATCCCGCCAGCGCGCGCTGA 973	
Db	3641 ACAAGCGCACCCAGAGATCTTGGAGAGTGCAGCTGGGCGATCCCGCCAGCGCGCGCTGA 3700	
Qy	974 AGAAG 1033	
Db	3701 AGAAG 3760	

Qy	1034 ACAGGAGCTTCCGCAAGTACACCGCTTTCACCATCCCGAGCATCAACAGAGACCCCG 1093	
Db	3761 ACAGGAGCTTCCGCAAGTACACCGCTTTCACCATCCCGAGCATCAACAGAGACCCCG 3820	
Qy	1094 GCATCCGCTACAGTACAAAGTGTGCTGCCAGAGGCTGGAAAGGCGAGCCCGAGCATCTTCC 1153	
Db	3821 GCATCCGCTACAGTACAAAGTGTGCTGCCAGAGGCTGGAAAGGCGAGCCCGAGCATCTTCC 3880	
Qy	1154 AGAGCAGATGACCAAGATCTGTGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGTCT 1213	
Db	3881 AGAGCAGATGACCAAGATCTGTGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGTCT 3940	
Qy	1214 ACCAGTACATGAGAGACCTGTGAGGAGGAGACCTTGGAGATCGGCGAGCAGCCGCGCCA 1273	
Db	3941 ACCA-----GGCCCCCTGTACGTGGGAGCGACCTTGGAGATCGGCGAGCAGCCGCGCCA 3994	
Qy	1274 AGATGAGGAGCTCGCAAGACCTGTGCTGGCTGGGGCTTCAACAACCCCGCAAGAGAGC 1333	
Db	3995 AGATGAGGAGCTCGCAAGACCTGTGCTGGCTGGGGCTTCAACAACCCCGCAAGAGAGC 4054	
Qy	1334 ACCAGAGGAGCCCGCTTCTGTGAGTGGGTACGAGCTGCACCCCGCAAGTGGAGCG 1393	
Db	4055 ACCAGAGGAGCCCGCTTCTGTGAGTGGGTACGAGCTGCACCCCGCAAGTGGAGCG 4108	
Qy	1394 TGCAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCTGTAAACGACATCCAGAGAGCTGG 1453	
Db	4109 TGCAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCTGTAAACGACATCCAGAGAGCTGG 4168	
Qy	1454 TGGGAGAGCTGAACTGGGCGAGCCAGATCTACCCCGGATCAAGTGGCGGAGCTGTGCA 1513	
Db	4169 TGGGAGAGCTGAACTGGGCGAGCCAGATCTACCCCGGATCAAGTGGCGGAGCTGTGCA 4228	
Qy	1514 AGCTGCTCGCGCGCGCCAAAGCCCTGACCGACATCGTGCCTTGAACGAGGCGCGAGC 1573	
Db	4229 AGCTGCTCGCGCGCGCCAAAGCCCTGACCGACATCGTGCCTTGAACGAGGCGCGAGC 4288	
Qy	1574 TGGAGTGGCGGAGAAACCGCGAGATCTGTGCGGAGCCGCTGCAGCGGCTGTACTACGAC 1633	
Db	4289 TGGAGTGGCGGAGAAACCGCGAGATCTGTGCGGAGCCGCTGCAGCGGCTGTACTACGAC 4348	
Qy	1634 CCAGCAAGAGCTGTGTGCGCGAGATCCAGAGAGAGGCGCACGACCTGTGACCTTACCAGA 1693	
Db	4349 CCAGCAAGAGCTGTGTGCGCGAGATCCAGAGAGAGGCGCACGACCTGTGACCTTACCAGA 4408	
Qy	1694 TCTACAGGAGCCCTTCAAGAACCTGAGACCGGCAAGTACGCCAAGATGGCGACCCGCC 1753	
Db	4409 TCTACAGGAGCCCTTCAAGAACCTGAGACCGGCAAGTACGCCAAGATGGCGACCCGCC 4468	
Qy	1754 ACACCAACGAGCTGAGCAGCTGACCGAGGCGCTGCAGAGATCGCATGGAGAGCATCG 1813	
Db	4469 ACACCAACGAGCTGAGCAGCTGACCGAGGCGCTGCAGAGATCGCATGGAGAGCATCG 4528	
Qy	1814 TGATCTGGGGAAGACCCCGCAAGTTCGCTTGGCCATCCAGAGAGAGACCTTGGAGACCT 1873	
Db	4529 TGATCTGGGGAAGACCCCGCAAGTTCGCTTGGCCATCCAGAGAGAGACCTTGGAGACCT 4588	
Qy	1874 GGTGAGCCGACTACTGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCC 1933	
Db	4589 GGTGAGCCGACTACTGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCC 4648	
Qy	1934 CCCTGTGAAGCTGTGTACCAAGTGGAGAGGAGCCCATCATCGCGCGCGAGACCTTCT 1993	
Db	4649 CCCTGTGAAGCTGTGTGTACCAAGTGGAGAGGAGCCCATCATCGCGCGCGAGACCTTCT 4708	
Qy	1994 AGGTGAGCGGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTTACGTGACCCGAGC 2053	
Db	4709 AGGTGAGCGGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTTACGTGACCCGAGC 4768	
Qy	2054 GGGGCGCGGAGAGATCGTGTGAGCTTGCAGAGACCAACCAAGAGAGCCGAGCTGAGG 2113	
Db	4769 GGGGCGCGGAGAGATCGTGTGAGCTTGCAGAGACCAACCAAGAGAGCCGAGCTGAGG 4828	





Qy 974 AGAAGAAGAGAGCGTGTGACCGTGTGACGTGGGCGACGCTACTTTCAGCGTCCCTGG 1033  
Db 2048 AGAAGAAGAGAGCGTGTGACCGTGTGACGTGGGCGAGCGCTACTTTCAGCGTCCCTGG 2107  
Qy 1034 ACAGGAGCTTCCGCAAGTACACCGCTTCCACCATCCCGACATCAACAGGAGACCCCG 1093  
Db 2108 ACAGGAGCTTCCGCAAGTACACCGCTTCCACCATCCCGACATCAACAGGAGACCCCG 2167  
Qy 1094 GCATCCGCTACCAAGTACAAAGTGTGCTGCCAGGCGTGAAGGGCAGCCCCAGCATCTTC 1153  
Db 2168 GCATCCGCTACCAAGTACAAAGTGTGCTGCCAGGCGTGAAGGGCAGCCCCAGCATCTTC 2227  
Qy 1154 AGACAGCATACCAAGATCTGTAGACCTTCCGCGCCGACCAACCGGAGATCGTATCT 1213  
Db 2228 AGACAGCATACCAAGATCTGTAGACCTTCCGCGCCGACCAACCGGAGATCGTATCT 2287  
Qy 1214 ACCAGTACATGGACGACCTGTACGTGGGCGACGACCTGGAGATCGGCGACACCGGCCA 1273  
Db 2288 ACCA-----GGCCCCCTGTACGTGGGCGACGACCTGGAGATCGGCGACCAACCGGCCA 2341  
Qy 1274 AGATCGAGAGCTGCGGAGACCTGTGCGCTGGGCGTTCAACACCCCGCAAGAGC 1333  
Db 2342 AGATCGAGAGCTGCGGAGACCTGTGCGCTGGGCGTTCAACACCCCGCAAGAGC 2401  
Qy 1334 ACCAGAGGAGCCCCCTTCTGTGGATGGGCTACGAGCTGCACCCCGCAAGTGGACCG 1393  
Db 2402 ACCAGAGGAGCCCCCTTCTGTGGATGGGCTACGAGCTGCACCCCGCAAGTGGACCG 2455  
Qy 1394 TGCAGCCCATCGAGTGCAGAGAGAGAGAGCTGCACCGTGAACGACATCCAGAGCTGG 1453  
Db 2456 TGCAGCCCATCGAGTGCAGAGAGAGAGAGCTGCACCGTGAACGACATCCAGAGCTGG 2515  
Qy 1454 TGGGCAAGCTGAACCTGGGCGGAGCAGATCTACCCCGGACATCAAGGTGCGCGAGCTGCA 1513  
Db 2516 TGGGCAAGCTGAACCTGGGCGGAGCAGATCTACCCCGGACATCAAGGTGCGCGAGCTGCA 2575  
Qy 1514 AGCTGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1573  
Db 2576 AGCTGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2635  
Qy 1574 TGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1633  
Db 2636 TGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2695  
Qy 1634 CCAGCAAGGACCTGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1693  
Db 2696 CCAGCAAGGACCTGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2755  
Qy 1694 TCTACAGGAGCCCTTCAAGAACCTGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1753  
Db 2756 TCTACAGGAGCCCTTCAAGAACCTGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2815  
Qy 1754 ACACCAACGAGCTGAAGCAGCTGACCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1813  
Db 2816 ACACCAACGAGCTGAAGCAGCTGACCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2875  
Qy 1814 TGATCTGGGCGAGACCCCGGAGTTCCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1873  
Db 2876 TGATCTGGGCGAGACCCCGGAGTTCCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2935  
Qy 1874 GGTGAGCCGACTTACTTGGGCGAGCCGCTGGATCCCGGAGTGGGAGTTGCTGAACACCCCGCC 1933  
Db 2936 GGTGAGCCGACTTACTTGGGCGAGCCGCTGGATCCCGGAGTGGGAGTTGCTGAACACCCCGCC 2995  
Qy 1934 CCCTGGTGAAGCTGTGGTACAGCTGGAGAGAGGCCCATCATCGGCGCGGAGACCTTCT 1993  
Db 2996 CCCTGGTGAAGCTGTGGTACAGCTGGAGAGAGGCCCATCATCGGCGCGGAGACCTTCT 3055  
Qy 1994 ACGTGGAGCGG 2053  
Db 3056 ACGTGGAGCGG 3115  
Qy 2054 GGGGCGGCGAGAGATCTGTAGCTGTGACCGGAGACCAACCAAGAGACCGAGCTGCAGG 2113

Db 3116 GGGGCGGCGAGAGATCTGTAGCTGTGACCGGAGACCAACCGAGACCGAGCTGCAGG 3175  
Qy 2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCGAGCGAGGTGAACATCGTGACCGAGAGCCAGT 2173  
Db 3176 CCATCCAGCTGGCCCTGCAGGACAGCGGCGAGCGAGGTGAACATCGTGACCGAGAGCCAGT 3235  
Qy 2174 AGCCCTGGGCGATCATCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2233  
Db 3236 AGCCCTGGGCGATCATCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3295  
Qy 2234 TCATCGAGCAGCTGATCAAGAGGAGAGAGGTGTACTGAGCTGGGTGCCCGCCACCAAGG 2293  
Db 3296 TCATCGAGCAGCTGATCAAGAGGAGAGAGGTGTACTGAGCTGGGTGCCCGCCACCAAGG 3355  
Qy 2294 GCATCGGCGGCAACGAGCAGATCCAGAGCTGGTGAGCAAGGGCATCGGCAAGGTGCTGT 2353  
Db 3356 GCATCGGCGGCAACGAGCAGATCCAGAGCTGGTGAGCAAGGGCATCGGCAAGGTGCTGT 3415  
Qy 2354 TCCTGAGCGGATCGATGCGGCGCATCGTGATCTACAGTACATGAGACGACCTGTACGTTG 2413  
Db 3416 TCCTGAGCGGATCGATGCGGCGCATCGTGATCTACAGTACATGAGACGACCTGTACGTTG 3475  
Qy 2414 GCAGCGGCGGCGCTAGGATCGATTAAGCTTCCCGGGGCTAGCACCGGT 2463  
Db 3476 GCAGCGGCGGCGCTAGGATCGATTAAGCTTCCCGGGGCTAGCACCGGT 3525

Search completed: April 10, 2004, 07:32:20  
Job time : 632.513 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 06:48:32 ; Search time 118.622 Seconds  
(without alignments)  
11550.795 Million cell updates/sec

Title: US-09-610-313-30

Perfect score: 2469

Sequence: 1 GCGAGCGCACCATGGCGCA.....GGGtagcaccgtgaattc 2469

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/prodata/2/ina/ECTUS\_COMB.seq:\*
- 6: /cgn2\_6/prodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2052	83.1	2312	4	US-09-475-515-84
2	2025.2	82.0	2306	4	US-09-475-515-82
3	1998.4	80.9	2300	4	US-09-475-515-83
4	1969.4	79.8	4319	4	US-09-475-515-6
5	1858	75.3	2305	4	US-09-475-515-80
6	1831.2	74.2	2299	4	US-09-475-515-81
7	1878.6	68.0	4307	4	US-09-552-950-2
8	1851.4	66.9	9772	4	US-09-552-950-5
9	1592.8	64.5	8366	4	US-09-872-733A-6
10	1557.4	63.1	4338	4	US-09-872-733A-1
11	1427.2	49.7	9010	4	US-09-184-418C-8
12	1196.2	48.4	8972	4	US-09-184-418C-9
13	1190.6	48.2	2467	4	US-09-872-733A-3
14	1189	48.2	8959	4	US-09-184-418C-11
15	1165.6	47.2	8992	4	US-09-184-418C-4
16	1132	45.8	2601	3	US-09-117-217-7
17	1132	45.8	2601	3	US-09-117-217-9
18	1132	45.8	2601	3	US-09-117-217-11
19	1132	45.8	2601	3	US-09-117-217-13
20	1132	45.8	2601	4	US-09-735-487-7
21	1132	45.8	2601	4	US-09-735-487-9
22	1132	45.8	2601	4	US-09-735-487-11
23	1132	45.8	2601	4	US-09-735-487-13
24	1132	45.8	4307	4	US-09-552-950-1
25	1132	45.8	9719	4	US-09-700-304-1
26	1128.8	45.7	9050	4	US-09-184-418C-7
27	1125.6	45.6	7399	2	US-08-418-848A-9

28 1125.6 45.6 9709 2 US-08-188-583-5 Sequence 5, Appli  
29 1125.6 45.6 9709 3 US-08-388-353-1 Sequence 1, Appli  
30 1125.6 45.6 9709 3 US-08-488-551B-1 Sequence 1, Appli  
31 1125.6 45.6 9709 4 US-09-309-572-15 Sequence 15, Appli  
32 1125.6 45.6 9709 4 US-09-718-096-15 Sequence 15, Appli  
33 1125.6 45.6 12479 4 US-09-318-138-13 Sequence 13, Appli  
34 1125.6 45.6 12494 3 US-08-935-312-13 Sequence 33, Appli  
35 1125.6 45.6 12494 3 US-08-848-760B-33 Sequence 35, Appli  
36 1125.6 45.6 15581 3 US-08-646-538-35 Sequence 35, Appli  
37 1125.6 45.6 15581 3 US-09-503-222-35 Sequence 35, Appli  
38 1122.8 45.5 3000 4 US-09-184-418C-74 Sequence 74, Appli  
39 1121 45.4 8968 4 US-09-184-418C-1 Sequence 1, Appli  
40 1116 45.2 4313 4 US-09-475-515-3 Sequence 3, Appli  
41 1116 45.2 9737 4 US-08-944-449-7 Sequence 7, Appli  
42 1116 45.2 9737 4 US-09-353-362-7 Sequence 6, Appli  
43 1108 44.9 8954 4 US-09-184-418C-6 Sequence 83, Appli  
44 1101.2 44.6 3017 4 US-09-184-418C-83 Sequence 101, App  
45 1099 44.5 3011 4 US-09-184-418C-101

#### ALIGNMENTS

##### RESULT 1

US-09-475-515-84

; Sequence 84, Application US/09475515A

; Patent No. 6602705

; GENERAL INFORMATION:

; APPLICANT: BARNETT, Susan

; APPLICANT: ZUR MEGEDE, Jan

; APPLICANT: SRIVASTAVA, Indresh

; APPLICANT: LIAN, Ying

; APPLICANT: HARTOG, Karin

; APPLICANT: LIU, Hong

; APPLICANT: GREER, Catherine

; APPLICANT: SELBY, Mark

; APPLICANT: WALKER, Christopher

; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION

; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES

; FILE REFERENCE: 1621.002

; CURRENT APPLICATION NUMBER: US/09/475,515A

; CURRENT FILING DATE: 1999-12-30

; NUMBER OF SEQ ID NOS: 90

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 84

; LENGTH: 2312

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:

; OTHER INFORMATION: FS (-).protmod.RTopt (+)

US-09-475-515-84

Query Match 83.1%; Score 2052; DB 4; Length 2312;  
Best Local Similarity 93.6%; Pred. No. 0;  
Matches 2165; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

QY 170 GCGCAAGGAGGGCCACCATGAAGGACTGCAAGCGCGCGAGGCAACTTCTTCGCG 229  
Db 1 GCGCGCGCAAGGACCAATGAAGATTGCACTGAGACAGGCTAATTTCTTCGCG 60  
QY 230 AGGACCTGGGCTTCCCGGAGGAGGCGCGGAGTTCCCGAGGACAGAACCGGCCA 289  
Db 61 AGGACCTGGGCTTCTCGAGGCAAGCGCGGAGTTTCAGCAGGACAGACCGGCCA 120  
QY 290 ACAGCCCAACAGCGGAGCTGCAGGTGCGCGCGG-----ACAACCCCGCAGGAGG 343  
Db 121 ACAGCCCAACCGCGGAGCTGCAGGTGTTGGGGCGGAGACACAGCTTGGCGAGG 180  
QY 344 CCGCGCGGAGCGCCAGGCGACCCCTG-----ACTTCCCGCCAGATCACCTGTGGCAGC 397  
Db 181 CCGCGCGGAGCGCGGAGGACCGCTGAGCTTAACCTTCCCGCCAGATCACCTGTGGCAGC 240

398 G C C C C T G G T G A G C A T C A A G G T G G G C G C C A G A T C A A G G A G G C C T G C T G G A C A C G G G C G 457  
400 G C C C C T G G T G A C A T C A G A T C G G C G C C A G C T C A A G G A G G C C T G C T G G A C A C G G G C G 300  
458 C G A C G A C A C C G T G C T G G A G A T A G A C C T G C C C G C A A G T G G A A G C C C A G A T A T C G 517  
460 C G A C G A C A C C G T G C T G G A G A T A G A C C T G C C C G C A A G T G G A A G C C C A G A T A T C G 360  
518 G C G C A T C G G C G G T T C A T C A A G T G G C C A G T A C C A G A T C C T G A T C G A G A T C G G 577  
520 G C G G A T C G G G G C T T C A T C A A G T G G C C A G T A C C A G A T C C T G A T C G A G A T C G G 420  
578 G C A A A G G C A T C G G C A C C G T G C T G A T C G G C C C A C C C C G T G A A C A T A T C G G C C G A 637  
600 G C C A A A G G C A T C G G C A C C G T G C T G A T C G G C C C A C C C C G T G A A C A T A T C G G C C G A 480  
638 A C A T G C T G A C C A G C T G G G T G C A C C C T G A C T T C C C A T C A G C C C A T C G A G A C C G T G C 697  
640 A C T G C T G A C C A G A T C G G T G C A C C C T G A C T T C C C A T C A G C C C A T C G A G A C C G T G C 540  
698 C C G T A A G C T G A A C C C G G C A T G G A C C C C C A A G T G A A G C A G T G G C C C C T G A C C G A G G 757  
700 C C G T A A G C T G A A C C C G G G A T G G A C C C C C A A G T C A A G C A G T G G C C C C T G A C C G A G G 600  
758 A G A G A T C A A G C C C T G A C C C C A T C T G C G A G G A T G G A A G A G G G C A A G A T C A C C A 817  
760 A G A A G A T C A A G C C C T G T G A G A T C T G A C C G A G A T G G A A G A G G G C A A G A T C A G A 660  
818 A G A T C G C C C C G A A C C C T A C A C A C C C C G T G T T G C C A T C A A G A A G A G A C A G A 877  
820 A G A T C G C C C C G A A C C C C T A C A C A C C C C G T G T T G C C A T C A A G A A G A G A C A G A 720  
878 C C A G T G G C G A G C T G T G A C T T C C G A G T G A C A G C A C C C A G C A C C C A G A C T T C T G G 937  
720 C C A G T G G C G A G C T G T G A C T T C C G A G T G A A C A A G C A C C C A G A C T T C T G G 780  
938 A G T G C A G T G G C A T C C C C A C C C C G C C T G A A A A G A A G A G A G C G T A C C G T G C 937  
780 A G T G C A G T G G C A T C C C C A C C C C G C C T G A A A A G A A G A G A G C G T A C C G T G C 840  
998 T G A G A G T G G G C A G C C T A C T T A G C G T G C C C T G G A C A G A G A C T T C C G A A G T A C C G 1057  
840 T G A G A G T G G G C A G C C T A C T T A G C G T G C C C T G G A C A G A G A C T T C C G A A G T A C C G 900  
1058 C C T T C A C C A T C C C C A G A T C A A C A A G A G A C C C C G C A T C C G T A C C A G T A C A A G T G C 1117  
900 C C T T C A C C A T C C C C A G A T C A A C A A G A G A C C C C G C A T C C G T A C C A G T A C A A G T G C 960  
1118 T G C C C A G G C T G A A G G C A G C C C A G A T C T T C C A G A G A G A T G A C C A A G A T C C T G G 1177  
960 T G C C C A G G C T G A A G G C A G C C C A G A T C T T C C A G A G A G A T G A C C A A G A T C C T G G 1020  
1178 A G C C T T C C G C C G C A A C C C G A G A T C G T G A T C T A C C A G T A C A T G A C G A C C T G T A C G 1237  
1020 A G C C T T C C G A G A G A C C C C A C A T C T G A T C T A C C A G T A C A T G A C G A C C T G T A C G 1080  
1238 T G G C A G C G A C C T G A G A T C G G C A G A C C C G C C A A G A T G A G A G C T G C G A A G C C 1297  
1080 T G G C A G C G A C C T G A G A T C G G C A G A C C C G C C A A G A T G A G A G C T G C G C A G C C 1140  
1298 T G C T G C G T G G G C T T C A C C C C G A C A A G A G C C C A G A G A G C C C C C T T C T G T 1357  
1140 T G C T G C G T G G G C T T C A C C C C G A C A A G A G C C C A G A G A G C C C C C T T C T G T 1200  
1358 G G A T G G G T A C A G T G C A C C C G A C A A G T G A C C G T G A C C C A T C G A G T G C C C G A 1417  
1200 G G A T G G G T A C A G T G C A C C C G A C A A G T G A C C G T G A C C C A T C A T G T G C C C G A 1260  
1418 A G G A G A G T G A C C G T G A C C A G A C T C C A G A G C T G T G G C A A G T G A C T G G C C A G C C 1477  
1260 A G G A G A G T G A C C G T G A C C A G A C T C C A G A G C T G T G G C A A G T G A C T G G C C A G C C 1320  
1478 A G A T C A C C C C G C A T A A G T G G C C A G C T G T G C A A G C T G C T G C G G C C C A A G C C C 1537

1321 A G A T T A C G C C G G C A T C A A G T G A A G C A G C T G C A A G C T G C T G C G C C A C C A A G C C C 1380  
1538 T G A C C G A C A T C G T C C C C T G A C C G A G A G C C C A G A C T G A G C T G G C C G A A A C C G G G A G A 1597  
1381 T G A C C G A G T G A T C C C C C T G A C C G A G A G C C C A G A C T G A G C T G G C C G A A A C C G G G A G A 1440  
1598 T C C T G C G C A C C C G T G C A C G G G T G A C T A C A C C C C A G A A A G A C C T G T G G C C G A G A 1657  
1441 T C C T G A A G A G C C C G T G C A G A G T G A C T A C G A C C C C A G A A A G A C C T G T G G C C G A G A 1500  
1658 T C C A A A G C A G G C C A C A C C A G T G A C C T A C A G A T C T A C C A G A G C C C T C A A A A A C C 1717  
1501 T C C A A A G C A G G C C A G C C A G T G A C C T A C A G A T C T A C C A G A G C C C T C A A A A C C 1560  
1718 T G A A A C C G G C A A G T A C C C A A G A T G C G A C C C C A C A C A A C A G A C G T G A A G A G T G A 1777  
1561 T G A A A C C G G C A A G T A C C C C A T G C G C G C C C A C A C A A C A G A C G T G A A G A G T G A 1620  
1778 C G A G C C C T G C A A A G A T C C C A T G G A G A C A C T G T A T C T G G G C A A G A C C C C A A G T 1837  
1621 C C G A G C C G T C A A A G T G A C C C A G A G A C T G G A G A C T G T G A T C T G G G C A A G A T C C C C A A G T 1680  
1838 T C C G C C C C C A T C C A A A G A G A C C T G G A G A C C T G T G A C C G A C T A C T G G A G C C A 1897  
1681 T C A A G C T C C C A T C C A A A G A G A C C T G G A G A C C T G T G A T G G A G T A C T G G C A G C C A 1740  
1898 C T G A T C C C G A G T G G A G T T C G T G A C A C C C C C C C C C T G T G A A C T G T G T A C C A G C 1957  
1741 C C T G A T C C C G A G T G G A G T T C G T G A C A C C C C C C C C C C T G T G A A C T G T G T A C C A G C 1800  
1958 T G G A A A G A G C C C A T C A T C G G C C C C A G A C C T T A C T G T G A C C G C C C C C A A C C G C G 2017  
1801 T G G A A A G A G C C C A T C G T G G C C C C A G A C C T T A C T G T G A C C G C C C C C A A C C G C G 1860  
2018 A G A C A A A T C G G C A A G C C C G C T A C T G A C C C A C C G G G C C G C A G A A G A T C G T G A C C 2077  
1861 A G A C A A G C T G G C A A G C C C G C T A C T G A C C C A C C G G G C C G C A G A A G T G T G A G C A 1920  
2078 T G A C C G A C A C C A C A C C A G A A G C C G A G T G C A G G C C A T C C A G C T G G C C C T G C A G A C A 2137  
1921 T G C C C G A C A C C A C C A C C A G A A C C G A G C T G C A G G C C A T C C A C C T G C C C T G C A G A C A 1980  
2138 G C G C A G C A G T G A A C A T C G T G A C C G A C C A G T A C C C C T G G G A T C A T C C A G C C C 2197  
1981 G C G C C T G A G T G A A C A T C G T G A C C A G C A G T A C C C C T G G G C A T C A T C C A G C C C 2040  
2198 A C C C G A C A A G C A G A G C A G C A G T G G T G A A C C A G A T C A T C G A C A G C T G A T C A A G A G G 2257  
2041 A C C C G A C A A G C A G A G C A G C A G T G G T G A C C A G A T C A T C G A C A G C T G A T C A A G A G G 2100  
2258 A G A A G T G T A C C T G A G T G G T G C C C C A C A A G G G A T C G G C G G C A C A G A G A T C G 2317  
2101 A G A A G T G T A C C T G C C T G G G T G C C C C C A A A G G G C A T C G G C G G C A C A G A G A G T G G 2160  
2318 A C A G C T G T G A G A A G G C A T C C G A A G T G C T T C C T G A C G G C A T C C A T G C G G C A 2377  
2161 A C A G C T G T G A G C C C G G C A T C C C A A G T G C T T C C T G A A C G G C A T C A T G C G G C A 2220  
2378 T C G T G A T C T A C C A G T A C A T G G A C A C C T G T A C T G G G C A G C G G C C C T A G G A T C G A T T 2437  
2221 T C G T G A T C T A C C A G T A C A T G G A C A C C T G T A C T G G G C A G C G G C C C T A G G A T C G A T T 2280  
2438 A A A G C T T C C C G G G C T A G C C C G T G A T T C 2469  
2281 A A A G C T T C C C G G G C T A G C C C G T G A T T C 2312

## RESULT 2

US-09-475-515-82  
; Sequence 82, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:

```

; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: WALKER, Mark
; APPLICANT: SELBY, Mark
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 82
; LENGTH: 2306
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.RTcpt.YM
US-09-475-515-82

Query Match      82.0%; Score 2025.2; DB 4; Length 2306;
Best Local Similarity 93.3%; Pred. No. 0;
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

QY 170 GCGGCAAGAGGCGCCACCAAGATGAAGAGTGCACCGAGCGCGAGCCCAACTTCTTCGCG 229
DB 1 GCGGCGCGAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAATTTCTTCGCG 60

QY 230 AGAGCTGCTTCCCGGAGGCGAGGCGCGAGTTCACCGAGAGAGAGCGCGCA 289
DB 61 AGAGCTGCTTCTGCGAGGCGAGGCGCGAGTTCACCGAGAGAGAGCGCGCA 120

QY 290 ACAGCGCCACCGCGCGAGTGCAGTGCAGGCGCGGCGG-----ACAAACCGCGCGAGG 343
DB 121 ACAGCGCCACCGCGCGAGTGCAGTGCAGGCGCGGCGGAGAGCAACAGCTGAGCGAGG 180

QY 344 CCGGCGCGAGGCGCGAGGCGCGCGG-----AACTTCCCGAGATCACTTTCGCGAGC 397
DB 181 CCGGCGCGAGGCGCGAGGCGCGCGGTTCACTTCCCGAGATCACTTTCGCGAGC 240

QY 398 GCGGCGCGAGGCGCGAGGCGCGCGAGTCAAGAGGCGCGCGAGTCAAGAGGCGCGCGG 457
DB 241 GCGGCGCGAGGCGCGAGGCGCGCGAGTCAAGAGGCGCGCGAGTCAAGAGGCGCGCGG 300

QY 458 CCGAGCACACCGTGTGAGGAGATGAGCTGCGCGCGAGTGAAGCGCGCGAGGATGATCG 517
DB 301 CCGAGCACACCGTGTGAGGAGATGAGCTGCGCGCGAGTGAAGCGCGCGAGGATGATCG 360

QY 518 GCGGCGCGAGGCGCGAGGCGCGCGAGTCAAGAGGCGCGCGAGTCAAGAGGCGCGCGG 577
DB 361 GCGGCGCGAGGCGCGAGGCGCGCGAGTCAAGAGGCGCGCGAGTCAAGAGGCGCGCGG 420

QY 578 GCAAGAGCGCATCGGCGCGAGTGCAGTGCAGCGCGCGAGTGAAGCGCGCGAGTGAAGCG 637
DB 421 GCAAGAGCGCATCGGCGCGAGTGCAGTGCAGCGCGCGAGTGAAGCGCGCGAGTGAAGCG 480

QY 638 ACATGCTGACCGCGCGAGTGCAGTGCAGCGCGCGAGTCAAGAGGCGCGCGAGTCAAGAG 697
DB 481 ACATGCTGACCGCGCGAGTGCAGTGCAGCGCGCGAGTCAAGAGGCGCGCGAGTCAAGAG 540

QY 698 CCGTGAAGTGAAGCGCGCGAGTGAAGCGCGCGAGTGAAGCGCGCGAGTGAAGCGCGCGAG 757
DB 541 CCGTGAAGTGAAGCGCGCGAGTGAAGCGCGCGAGTGAAGCGCGCGAGTGAAGCGCGCGAG 600

QY 758 AGAAGATCAAGCGCGCGAGTGAAGCGCGCGAGTGAAGCGCGCGAGTGAAGCGCGCGAG 817
DB 601 AGAAGATCAAGCGCGCGAGTGAAGCGCGCGAGTGAAGCGCGCGAGTGAAGCGCGCGAG 660

```

```

QY 818 AGATCGGCGCGGAGAAACCCCTACAAACACCCCGGTGTTCGCCATCAAGAAAGGACAGCA 877
DB 661 AGATCGGCGCGGAGAAACCCCTACAAACACCCCGGTGTTCGCCATCAAGAAAGGACAGCA 720

QY 878 CCAAGTGGCGGAGAGCTGGTGGACTTCGCGGAGTGAAGAGCGACCCAGGACTTCTGGG 937
DB 721 CCAAGTGGCGGAGAGCTGGTGGACTTCGCGGAGTGAAGAGCGACCCAGGACTTCTGGG 780

QY 938 AGGTGAGCTGGGCGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 997
DB 781 AGGTGAGCTGGGCGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840

QY 998 TGAAGTGGCGGAGCGGCTTACTTCAGCGTGCCTTCAGCGAGGACTTCCCGAGTACACCG 1057
DB 841 TGAAGTGGCGGAGCGGCTTACTTCAGCGTGCCTTCAGCGAGGACTTCCCGAGTACACCG 900

QY 1058 CTTTCAACATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1117
DB 901 CTTTCAACATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960

QY 1118 TGCCCGAGGCTGGAAAGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1177
DB 961 TGCCCGAGGCTGGAAAGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020

QY 1178 AGCCCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1237
DB 1021 AGCCCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1074

QY 1238 TGCGGAGCGAGCTGGAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1297
DB 1075 TGCGGAGCGAGCTGGAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1134

QY 1298 TGCTGCGCTGGGCGCTTCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1357
DB 1135 TGCTGCGCTGGGCGCTTCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1194

QY 1358 GATGCGCTACGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1417
DB 1195 GATGCGCTACGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1254

QY 1418 AGAGAGCTGGAGCGTGAAGAGCATCCAGAGCTGGTGGGCGAGCTGAAGCTGGCGCGCG 1477
DB 1255 AGAGAGCTGGAGCGTGAAGAGCATCCAGAGCTGGTGGGCGAGCTGAAGCTGGCGCGCG 1314

QY 1478 AGATCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1537
DB 1315 AGATCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1374

QY 1538 TGACCGAGATCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1597
DB 1375 TGACCGAGGCTGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1434

QY 1598 TCCTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1657
DB 1435 TCCTGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1494

QY 1658 TCAGAGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1717
DB 1495 TCAGAGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1554

QY 1718 TGAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1777
DB 1555 TGAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1614

QY 1778 CCGAGCGCGTGCAGAGAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1837
DB 1615 CCGAGCGCGTGCAGAGAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1674

QY 1838 TCCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1897
DB 1675 TCAAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1734

QY 1898 CTTGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1957

```

```

1735 CCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTACAGC 1794
1958 TGGAGAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGACGGCGCGCAACCGCG 2017
1795 TGGAGAGGAGCCCATCTGTGGCGCGAGACCTTCTACGTGGAGCGCGCGCAACCGCG 1854
2018 AGACCAAGATGGCAAGCGCGGTACGTGACCGACCGCGCGCGCGCAAGATCGTAGCC 2077
1855 AGACCAAGTGGCAAGCGCGGTACGTGACCGACCGCGCGCGCGCAAGATCGTAGCA 1914
2078 TGACCGAGACCAACCAAGAGACCGAGCTGACGAGCCATCCAGCTGGCGCCCTCAGGACA 2137
1915 TCGCGGACACCAACCAAGAGACCGAGCTGACGAGCCATCCAGCTGGCGCCCTCAGGACA 1974
2138 GCGCAGCGAGGTGAACATCGTGACCGACCGACCGAGTACGCGCCCTGGCGCATCATCAGGCC 2197
1975 GCGCGCTGGAGTGAACATCGTGACCGACCGAGTACGCGCCCTGGCGCATCATCAGGCC 2034
2198 AGCCGACAGAGCGAGCGAGTGGTGAACCGAGTACGAGCATTCGAGCAGCTGATCAAGAGG 2257
2035 AGCCGACAGAGCGAGCGAGTGGTGAACCGAGTACGAGCATTCGAGCAGCTGATCAAGAGG 2094
2258 AGAAGGTGTACTCTGAGTGGTGGTGGCGCCCGCCACAGAGGCAATCGGCGCAACGAGCATCG 2317
2095 AGAAGGTGTACTCTGAGTGGTGGTGGCGCCCGCCACAGAGGCAATCGGCGCAACGAGCATCG 2154
2318 ACAGCTGGTGAACAGGCGATCGCAAGTGTCTTCTGGACGCGATCGATGGCGGCA 2377
2155 ACAAGCTGGTGAAGCGCGGCAATCGCAAGTGTCTTCTGGACGCGATCGATGGCGGCA 2214
2378 TCAGTGTACTACAGTACATGACGACCTGTACGTGGCGCGCGCGCCCTAGGATCGATT 2437
2215 TCAGTGTACTACAGTACATGACGACCTGTACGTGGCGCGCGCGCGCCCTAGGATCGATT 2274
2438 AAAAGCTTCCCGGGGTAGCAACCGGTGATTTC 2469
2275 AAAAGCTTCCCGGGGTAGCAACCGGTGATTTC 2306

```

RESULT 3

```

; Sequence 83, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 83
; LENGTH: 2300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: PS(-).protmod.RTopt.YMMW
US-09-475-515-83

```

Query Match 80.9%; Score 1998.4; DB 4; Length 2300;  
Best Local Similarity 92.9%; Pred. No. 0;  
Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;

```

170 GCGCAAGAGGAGCGCCACAGATGAAGATGACCGAGCGCCAGGCCAACATCTTCTCCGCG 229
1 GCGGCGCGAAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAAATTTCTTCCGCG 60
230 AGGACCTGGCCCTTCCCGAGGCAAGCGCGAGTTCCCGAGCGAGCAAGAACCGCGCCA 289
61 AGGACCTGGCCCTTCTGAGGGCAAGCGCGAGTTCCCGAGCGAGCAAGAACCGCGCCA 120
230 ACAGCCCCACAGCGCGAGCTGAGGTGGCGCGC-----ACAAACCGCGAGCGAGG 343
121 ACAGCCCCACCGCGCGAGCTGAGGTGGCGCGCGAGAAACACAGCTGAGCGAGG 180
344 CGGCGCGAGCGCGAGCGCGACCGCTG-----ACTTCCCGAGATCACCTGTGGCAGC 397
181 CGGCGCGAGCGCGAGCGCGACCGTGAAGTTCACTTCCCGAGATCACCTGTGGCAGC 240
398 GCGCCCTGGTGAAGCAI CAAGGTGGCGCGCGAGATCAAGAGGCGCCCTCTGAGACACCGCG 457
241 GCGCCCTGGTGAAGCAI CAAGGTGGCGCGCGAGCTCAAGAGGCGCTCTGAGACACCGCG 300
458 CGGAGCACACCGTGTGAGAGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCG 517
301 CGGAGCACACCGTGTGAGAGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCG 360
518 GCGGATCGCGCGCTTCAATCAAGGTGGCGCGAGTACGACAGATCCTGATCGAGATCTGCG 577
361 GCGGATCGCGCGCTTCAATCAAGGTGGCGCGAGTACGACAGATCCTGATCGAGATCTGCG 420
578 GCAAGAGCGCATCGCAACCGTGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 637
421 GCAAGAGCGCATCGCAACCGTGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
638 ACATGCTGACCGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 697
481 ACCTGCTGACCGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
698 CGGTGAAGCTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 757
541 CGGTGAAGCTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
758 AGAAGATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 817
601 AGAAGATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
818 AGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 877
661 AGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
878 CCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 937
721 CCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
938 AGGTGACGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 997
781 AGGTGACGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
998 TGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1057
841 TGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
1058 CCTTCAACCATCCCGAGCATCAACAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1117
901 CCTTCAACCATCCCGAGCATCAACAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
1118 TGGCGCGAGGTGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1177
961 TGGCGCGAGGTGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
1178 AGCCCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1237
1021 AGCCCTTCCGCGAGCAGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1074

```

QY 1238 TGGGAGCGACCTGGAGATCGGCGACGACCGGCCAAGATCGAGAGCTGGCGAAGCACC 1297  
Db 1075 TGGGAGCGACCTGGAGATCGGCGACGACCGGCCAAGATCGAGAGCTGGCGAAGCACC 1134  
QY 1298 TCGTGGCTGGGCTTCCACCCCGGACCAAGACGACCAAGAGAGCCCGCTTCTGT 1357  
Db 1135 TCGTGGCTGGGCTTCCACCCCGGACCAAGACGACCAAGAGAGCCCGCTTCTGT 1194  
QY 1358 GGATGGGCTACGAGGTGACCCCGGACCAAGTGCAGCCCATCGAGTGGCCGAGA 1417  
Db 1195 CCAT-----CGAGCTGCACCCCGGACCAAGTGCAGCCCATCGAGTGGCCGAGA 1248  
QY 1418 AGGAGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGGCAAGCTGAACCTGGCGACCC 1477  
Db 1249 AGGAGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGGCAAGCTGAACCTGGCGACCC 1308  
QY 1478 AGATCTACCCCGGACCTCAAGGTGGCGAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCC 1537  
Db 1309 AGATCTACCCCGGACCTCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCC 1368  
QY 1538 TGACCGACATCGTGGCCCTGACCGAGAGGCGCGAGCTGGAGTGGCCCGAGAACCGCGAGA 1597  
Db 1369 TGACCGAGGTGATCCCGCTGACCGAGAGGCGCGAGCTGGAGTGGCCCGAGAACCGCGAGA 1428  
QY 1598 TCGTGGCGAGCCGTGACCGCGCTGTACTAGACCCGACGCAAGGACCTGGTGGCCGAGA 1657  
Db 1429 TCGTGAAGAGCCGTGACCGAGGTGTACTAGACCCGACGCAAGGACCTGGTGGCCGAGA 1488  
QY 1658 TCAGAAGCAGGCGCACGACCACTGAGCTGTACAGATCTTACAGAGGCCCTTCAAGAAC 1747  
Db 1489 TCAGAAGCAGGCGCACGAGGCCAGTGGACCTTACAGATCTTACAGAGGCCCTTCAAGAAC 1548  
QY 1718 TGAAGACCGGCAAGTACGCGCAGATGCGCAGCCGCGCACCAAGACGAGTGAAGCGCTGA 1777  
Db 1549 TGAAGACCGGCAAGTACGCGCAGATGCGCAGCCGCGCACCAAGACGAGTGAAGCGCTGA 1608  
QY 1778 CGAGGCGGTGACAGAGATCGCCATGGAGAGCATCTGTATCTGGGGCAAGACCCCAAGT 1837  
Db 1609 CGAGGCGGTGACAGAGTGGAGTGGAGAGCATCTGTATCTGGGGCAAGATCCCAAGT 1668  
QY 1838 TCGGCTGCGCATCCAGAGGAGACCTGGAGACCTGTGTGACCGACCTACTGGCAGGCCA 1897  
Db 1669 TCAAGCTGCGCATCCAGAGGAGACCTGGAGGCGCTGTGTGATGGAGTACTGGCAGGCCA 1728  
QY 1898 CTGTATCCCGAGTGGGAGTTCGTGAACACCCCGCGCGCTGTGTGAAGTGTGTACCAAGC 1957  
Db 1729 CTGTATCCCGAGTGGGAGTTCGTGAACACCCCGCGCGCTGTGTGAAGTGTGTACCAAGC 1788  
QY 1958 TGGAGAGAGCCCATCATCGGCGCGCGAGACCTTCTAGTGAACCGCGCGCGCGCAACCGCG 2017  
Db 1789 TGGAGAGAGGCCCATCTGTGGCGCGCGAGACCTTCTAGTGAACCGCGCGCGCGCAACCGCG 1848  
QY 2018 AGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGCGCGCGCGCGAGAGATCGTGAGCC 2077  
Db 1849 AGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGCGCGCGCGCGAGAGTGTGTAGCA 1908  
QY 2078 TGAACGAGACCAACCAAGAGAGACCGAGCTGACAGCCCATCAAGTGGCCCTGAGAGACA 2137  
Db 1909 TCGCGGACACCAACCAAGAGAGACCGAGCTGACAGCCCATCAAGTGGCCCTGAGAGACA 1968  
QY 2138 GGGGAGAGAGTGAACATCGTGACCGACGCGCTAGCCCTGGGATCATCGAGGCC 2197  
Db 1969 GCGGCTGGAGGTGAACATCGTGACCGACGCGCTAGCCCTGGGATCATCGAGGCC 2028  
QY 2198 AGCCGCAAGAGCGAGAGCGAGCTGGTGAACAGATCATCGAGCAGCTGATCAAGAGG 2257  
Db 2029 AGCCGCAAGAGCGAGAGCGAGCTGGTGAACAGATCATCGAGCAGCTGATCAAGAGG 2088  
QY 2258 AGAGGCTGACCTGAGCTGGTGGCGCGCGCACAGGGCATCGGCGGCAACAGAGATCG 2317  
Db 2089 AGAGGCTGACCTGAGCTGGTGGCGCGCGCACAGGGCATCGGCGGCAACAGAGATCG 2148  
QY 2318 ACAAGCTGTGAGCAAGGCGATCCCGAAGGTGCTGTTCTGAGCGGCAATCGATGGCGCA 2377

Db 2149 ACAAGCTGTGAGCGCGCATCCGCAAGGTGTGTTCTCTGAACGGCATCGATGGCGCA 2208  
QY 2378 TCGTGTATCTACAGTACATGGAGCACTGTACTGGGAGAGCGCGCGCTAGATCGATT 2437  
Db 2209 TCGTGTATCTACAGTACATGGAGCACTGTACTGGGAGAGCGCGCGCTAGATCGATT 2268  
QY 2438 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2469  
Db 2269 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2300  
  
RESULT 4  
US-09-475-515-6  
; Sequence 6, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEDEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE OF INVENTION: OF VIRUS-LIKE PARTICLES  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 4319  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: HIV-Gag-polymerase  
US-09-475-515-6  
  
Query Match 79.8%; Score 1969.4; DB 4; Length 4319;  
Best Local Similarity 90.1%; Pred. No. 0;  
Matches 2159; Conservative 0; Mismatches 221; Indels 16; Gaps 4;  
  
QY 14 TGGCCGAGGCGCATGAGCCAGGCCACCAGC---GCCAATCTCTGATGCGAGCGCAACT 70  
Db 1100 TGGCCGAGGCGCATGAGCCAGGTTGACGACCCCGCGCATCATGATGCGAGCGGCACT 1159  
QY 71 TCAAGGGCCCCAAGCGCATCATCAAGTGTTCATCTGCGGCAAGGAGGCCACATCGCCC 130  
Db 1160 TCGCAACCAAGCGGAAGACCGTCAAGTGTTCATCTGCGCAAGGAGGCCACACGCCA 1219  
QY 131 GCAATGCGCGCGCCCGCGCAAGAGGCTGTGGAAGTGGCGCAAGGAGGCCACCA 190  
Db 1220 GGAATCTCGCGCGCCCGCGCAAGAGGCTGTGCGCTGCGCGCGCGCAAGGAGCA 1279  
QY 191 TGAAGAGCTGACCGAGCGCCAGGCCAATCTTCTCGCGAGGACCTGGGCTTCCCGCAG 250  
Db 1280 TGAAGATTCATCTGAGACACAGGCTTA---TTTTAGGGAAGATCTGGCTTCTTACAG 1338  
QY 251 GCAAGCGCCCGAGTTCGCCAGCGAGAGAACCGCGGCCAACAGCCCAACAGCGCGAGC 310  
Db 1339 GGAAGCGCCAGGAATTTTCTTCAGAGCAGACAGAGCCCAACAGCCCAACAGAGAGC 1398  
QY 311 TGCAGGTGCGCGCG---ACAAACCCCGAGAGGAGCGCGCGCGCGAGCGCCAGGCA 364  
Db 1399 TGCAGGTGCGCGCGAGAGAGAAACAATCTCTCTCAGAACAGAGCGCGGATAGACAGNA 1458  
QY 365 -----CCCTGAATTTCCCGCAGATCACCTGTGGCAGCGCCCGCTGTGTGATCAAG 418  
Db 1459 CTGTATCTTAACTTCCCTCAGATCACTTTTGGCAAGACCCCTCGTCAAGTAAGA 1518

1499	TGCGCCAGCTGTGAAGTGTCTGCGCGCGCCAAAGGCTCCAGCGACATCTGTCGCCCTGA	1558	QY
2599	TGAAGCAGCTGTGAAGTGTCTGCGCGGCAACCAAGGCCCTGACCGAGTGATCCCCCTGA	2658	Db
1559	CCGAGGAGCCGAGCTGGAGTGTGCCCGAGAACCCGCGAGATCCTTCGGAGACCCCTGTCAACG	1618	QY
2659	CCGAGGAGCCGAGCTGGAGTGTGCCCGAGAACCCGCGAGATCCTTCGAAGAGAGCCCTGTCAACG	2718	Db
1619	CGGTGTACTACGACCCCGACGACGAGCACTGTGTGGCCGAGATCCAGAAAGCAGAGGCCACGACC	1678	QY
2719	AGGTGTACTACGACCCCGACGAGCACTGTGTGGCCGAGATCCAGAAAGCAGAGGCCACGAGCC	2778	Db
1679	AGTGGACCTACAGATCTTACGAGAGGCCCTTTCAAGAACCTGAAGACCCGGCAAGTACGCCA	1738	QY
2779	AGTGGACCTACAGATCTTACGAGAGGCCCTTTCAAGAACCTGAAGACCCGGCAAGTACGCCC	2838	Db
1739	AGATGCGCACCGCCACACCAACGACGTGAAGCAGCTGACCGAGGCCCTGCAGAAAGATCG	1798	QY
2839	GCATGCGCGCGGCCACACCAACGACGTGAAGCAGCTGACCGAGGCCCTGCAGAAAGTGA	2898	Db
1799	CCATGGAGAGCATCGTGATCTTGGGGCAAGACCCCGAAGTTCCGCTGCCCATCCAGAAGG	1858	QY
2899	GCACCGAGAGCATCGTGATCTTGGGGCAAGATCCCGAAGTTCAAGGTGCCCATCCAGAAGG	2958	Db
1859	AGACCTGGAGACCTGTGTGGACCGCACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGT	1918	QY
2959	AGACCTGGAGGCCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCGAGTGGGAGT	3018	Db
1919	TGCTGAACACCCGCCCTGTGTGAAGCTGTGGTACCAGCTGGAGAGAGGCCCATCATCG	1978	QY
3019	TGCTGAACACCCGCCCTGTGTGAAGCTGTGGTACCAGCTGGAGAGAGGCCCATCATCTGGG	3078	Db
1979	GGCGCCAGACCTTCTACGTGGACGGCGCGGCCAACCGCGAGACCAAGATCGCAAGGCCG	2038	QY
3079	GGCGCCAGACCTTCTACGTGGACGGCGCGGCCAACCGCGAGACCAAGCTGGCAAGGCCG	3138	Db
2039	GCTAGCTACCGACCGGGCGCGGAGAGATCGTGAGCCTGACCGAGACCCACCAACCGA	2098	QY
3139	GCTAGCTACCGACCGGGCGCGGAGAGTGGTAGCATCCGCGACACCAACCAACCGA	3198	Db
2099	AGACCGAGCTGCAGGCCATCCAGCTGGCGCCCTGCAGGACAGCGGCAGCGAGTGAACATCG	2158	QY
3199	AGACCGAGCTGCAGGCCATCCAGCTGGCGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCG	3258	Db
2159	TGACCGACAGCCAGTACGCCCTGGCGCATCTCCAGCGCCACGCCGACAGAGCGAGAGCG	2218	QY
3259	TGACCGACAGCCAGTACGCCCTGGCGCATCTCCAGCGCCACGCCGACAGAGCGAGAGCG	3318	Db
2219	AGCTGGTGAACCCAGATCATTCGAGCAGCTGATCAAGAGGAGAGGTGTACTTGAAGCTGGG	2278	QY
3319	AGCTGGTGAAGCCAGATCATTCGAGCAGCTGATCAAGAGGAGAGGTGTACTTGGCTGGG	3378	Db
2279	TGCCCCGCCACAAAGGGCATTCGCGCCGACGAGCAGATCGACAAAGCTGGTGAACAAGGGCA	2338	QY
3379	TGCCCCGCCACAAAGGGCATTCGCGCGCAACGAGCAGGTGGAACAGTGTGTGAGCGCCGCA	3438	Db
2339	TCCGCAAGGTGCTGTCTCTGGACGCATCGATGGCGGCATCTGTGATCTTACCAAGTA	2393	QY
3439	TCCGCAAGGTGCTGTCTCTGGAAACGCATCGCAACAGGCCACGAGAGGACACGAA	3493	Db

RESULT 5  
US-09-475-515-80  
; Sequence 80, Application US/09475151A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine



APPLICANT: SELBY, Mark  
APPLICANT: WALKER, Christopher  
TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
FILE REFERENCE: 1621.002  
CURRENT APPLICATION NUMBER: US/09/475,515A  
CURRENT FILING DATE: 1999-12-30  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 80  
LENGTH: 2305  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
OTHER INFORMATION: FS(+).prolnact.Rtopt.YM  
US-09-475-515-80

Query Match 75.3%; Score 1858; DB 4; Length 2305;  
Best Local Similarity 89.0%; Pred. No. 1.2e-298;  
Matches 2058; Conservative 0; Mismatches 235; Indels 19; Gaps 4;

QY 170 GCGGCAAGGAGGCGCCACAGATGAAGGACTGACCGAGCGCCAGGCCAACTTCTTCGCG 229  
DB 1 GCGGCGCGAGGACACCAATGAAGATTGCACTGAGAGACAGCTAATTTTT-AGGG 59

QY 230 AGGACTGCGCTTCCCGCCAGGCGCAAGCGCGAGTTCCCGAGCGAGCAACCGCGCCA 289  
DB 60 AAGATCTGCGCTTCTACAGGGAAGGCGCAGGGAATTTTCTTCAGAGCAGACAGAGCCA 119

QY 290 ACAGCCCGACAGCGCGAGCTGCGAGTGCAGCGCGG-----ACAAACCCCGCAGCGAGG 343  
DB 120 ACAGCCCGACAGAGAGCTTCAAGGTTGGGAGGAGAAACAACTCCCTCTCAGAA 179

QY 344 CCGGCGCGAGCGCCAGGCGCA-----CCCTGAATTCCTCCCGAGATCACTCTGGCAGC 397  
DB 180 CAGGAGCGCATAGACAGGAAGTGTATCTTTAACTTCCTTCAGATCACTCTTTGGCAAC 239

QY 398 GCGCCCTGTGAGCATCAAGTTGGCGGCGCAGATCAAGGAGCGCTGTGACACCGCGC 457  
DB 240 GACCCCTCGTCAATAAGATTCGGGGGCAACTCAAGGAGCGTGTCTCGATCAGGAG 299

QY 458 CGAGCACACCGTGTGGAGGAGATGAGCTTCGCGCGCAAGTGAAGCCCAAGATGATCG 517  
DB 300 CAGATGATACAGTATTAGAGAAATGAATTTCCAGGAAATGGAACCAAAATGATAG 359

QY 518 GCGGCAATCGCGGCTTCAATCAAGTTGCGCGCAGTACGACAGATCTGATCGAGATCTGCG 577  
DB 360 GCGGGAATCGGGGCTTCAATCAAGTTGAGGAGTACGACAGATCTGTAGAAATCTGTG 419

QY 578 GCAAGAGCGCATCGSCACCGTGTGATCGGCGCCACCGCTGAACATCATCGGCGCA 637  
DB 420 GACATGAAGCTATAGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGAAGAA 479

QY 638 ACATGCTGACCGAGTGGGCTGACCTGAACTTCCCGCATAGCCCGCATCGAGACCGTGC 697  
DB 480 ATCTGTGACCGAGATCGGCTGCACCTTGAATTCCTCCCATCAGCCCTATTGAGACGCTG 539

QY 698 CCGTGAAGCTGAAGCCCGCATGAGCGCCCAAGTGAAGCAGTGGCCCTGACCGAGG 757  
DB 540 CCGTGAAGTTGAAGCCCGGATGAGCGCCCAAGTCAAGTCAAGCAATGGCCATTCACCGAG 599

QY 758 AGAAGATCAAGGCGCTTACCGCATCTGCGAGGAGATGGAGAGGAGGCGCAAGATCACCA 817  
DB 600 AGAAGATCAAGGCGCTTGGTGGAGATCTGCAACCGAGATGGAGAGGAGGCGCAAGATCAGCA 659

QY 818 AGATCGGCGCGAGACCCCTCAACACCCCGCTTTCGCCATCAGAGAGGAGCAGCA 877  
DB 660 AGATCGGCGCGAGAACCCCTCAACACCCCGCTTTCGCCATCAGAGAGGAGCAGCA 719

QY 878 CCAAGTGGCGCAGCTGTGAGATTCGCGGAGCTGAACAGGCGCACCCAGGACTTCTGGG 937  
DB 720 CCAAGTGGCGCAGCTGTGAGATTCGCGGAGCTGAACAGGCGCACCCAGGACTTCTGGG 779

QY 938 AGGTGAGCTGGGCGATCCCGCACCGCGCGCTGAAGAAGAGAGCGGTGACCGTGC 997  
DB 780 AGGTGAGCTGGGCGATCCCGCACCGCGCGCTGAAGAAGAGAGCGGTGACCGTGC 839

QY 998 TGAACCTGGGCGAGCGCTTCTTACGCTGCCCTCGACGAGACTTCCGCAAGTACACG 1057  
DB 840 TGAACCTGGGCGAGCGCTTCTTACGCTGCCCTCGACGAGACTTCCGCAAGTACACG 899

QY 1058 CTTTCAACCATCCCGACGATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117  
DB 900 CTTTCAACCATCCCGACGATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 959

QY 1118 TGCCCGAGGCTGGAGAGGAGCGCGCGATCTTCCAGAGAGAGAGAGAGAGAGAGAG 1177  
DB 960 TGCCCGAGGCTGGAGAGGAGCGCGCGATCTTCCAGAGAGAGAGAGAGAGAGAGAG 1019

QY 1178 AGCCCTTCCGCGCGCGAGATCGTGTATCTTACAGTACATCGAGAGAGAGAGAGAG 1237  
DB 1020 AGCCCTTCCGCGAG 1073

QY 1238 TGCGCAGCGACCTGGAGATCGGCGAGCAGCGCGCGAGAGAGAGAGAGAGAGAGAG 1297  
DB 1074 TGCGCAGCGACCTGGAGATCGGCGAGCAGCGCGAGAGAGAGAGAGAGAGAGAG 1133

QY 1298 TGCTGCGCTGGGCTTACACACCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357  
DB 1134 TGCTGCGCTGGGCTTACACACCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1193

QY 1358 GGATGCGCTACGAGCTGCGACCGCGAGCAAGTGCAGCGCGAGAGAGAGAGAGAGAG 1417  
DB 1194 GGATGCGCTACGAGCTGCGACCGCGAGCAAGTGCAGCGCGAGAGAGAGAGAGAG 1253

QY 1418 AGAGAGCTGGACCGTGAAGAGATCCAGAGCTGGTGGGCAAGTGAAGTGGGCGAGC 1477  
DB 1254 AGAGAGCTGGACCGTGAAGAGATCCAGAGCTGGTGGGCAAGTGAAGTGGGCGAGC 1313

QY 1478 AGATCTACCGCGCGCATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGCGCGAGAG 1537  
DB 1314 AGATCTACCGCGCGCATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGCGCGAGAG 1373

QY 1538 TGACCGAGCATGTGCGCGCGTGAAGAGAGCGAGCTGGAGTGGCGCGAGAGAGAGAG 1597  
DB 1374 TGACCGAGCTGTGCGCGCGTGAAGAGAGCGAGCTGGAGTGGCGCGAGAGAGAGAG 1433

QY 1598 TCCTGCGCGAGCGCGTGAAGAGAGCTGTGCAAGCTGTGCAAGCTGTGCGCGCGAGAG 1657  
DB 1434 TCCTGAGAGAGCGCGTGAAGAGAGCTGTGCAAGCTGTGCAAGCTGTGCGCGCGAGAG 1493

QY 1658 TCAGAGAGCGGCGCGAGAGCTGTGCAAGCTGTGCAAGCTGTGCAAGCTGTGCAAGAG 1717  
DB 1494 TCAGAGAGCGGCGCGAGAGAGCTGTGCAAGCTGTGCAAGCTGTGCAAGAGAGAGAG 1553

QY 1718 TGAAGCGCGAG 1777  
DB 1554 TGAAGCGCGCGAG 1613

QY 1778 CCGAGCGCGTGAAG 1837  
DB 1614 CCGAGCGCGTGAAG 1673

QY 1838 TCAGCGCTGCGCGAG 1897  
DB 1674 TCAGCGCTGCGCGAG 1733

QY 1898 CTTGAGTCCCGAGTGGGAGTTCGTGAACACCCCGCGCGAGAGAGAGAGAGAGAGAG 1957  
DB 1734 CTTGAGTCCCGAGTGGGAGTTCGTGAACACCCCGCGCGAGAGAGAGAGAGAGAGAG 1793

QY 1958 TGGAG 2017  
DB 1794 TGGAG 1853



1358 GGATGGGCTACGAGTGTCAACCCCGACAAAGTGGACCGTGCAGCCCATCAGAGTGTCCCGAGA 1417  
1194 CCAT-----CGAGTGCACCCCGACAAAGTGGACCGTGCAGCCCATCAGTGTCCCGAGA 1247  
1418 AGGAGAGTGTGACCGTGTGAACGACATCCAGAAAGTGTGGGCAAGTGAACCTGGGCCAGCC 1477  
1248 AGGACAGTGTGACCGTGTGAACGACATCCAGAAAGTGTGGGCAAGTGAACCTGGGCCAGCC 1307  
1478 AGATCTACCCGGGACCAAGTGTGGCCAGCTGTGCAAGCTGTCTGGGGGCGCCCAAGGCC 1537  
1308 AGATCTACCCGGGACCAAGTGTGAAGCAAGCTGTGCAAGCTGTCTGGGGGCGCCCAAGGCC 1367  
1538 TGACCGACATCGTGTCCCTGTGACCGAGGAGCGGAGCTGGAGCTGGCGAGAACCGCCAGA 1597  
1368 TGACCGAGGTGTCTCCCTGTGACCGAGGAGCGGAGCTGGAGCTGGCGAGAACCGCCAGA 1427  
1598 TCTCCGGAGCCCTGTGACCGGCTGTACTACGACCCCGACCAAGCACTGTGTGCCCGAGA 1657  
1428 TCTGAAGAGGCGCTGTGACCGAGGTGTACTACGACCCCGACCAAGCACTGTGTGCCCGAGA 1487  
1658 TCCAGAACGCGGACGACGACGAGTGGAGCTTACAGATCTACAGAGGCCCTTCAAGAAC 1717  
1488 TCCAGAACGCGGACGAGGCGGAGTGGAGCTTACAGATCTACAGAGGCCCTTCAAGAAC 1547  
1718 TGAAGACCGGCAAGTACCGCAAGATGCGCACCGGCCACACCAACGACGCTGAAGCAGCTGA 1777  
1548 TGAAGACCGGCAAGTACCGCCGATGCGCGGCCACACCAACGACGCTGAAGCAGCTGA 1607  
1778 CGGAGCGGTGACAGAAATCGCCATGAGAGAGATCGTGTATCTGGGGCAAGACCCGCCAAGT 1837  
1608 CGGAGCGGTGACAGAAATGACACCGGAGAGATCGTGTATCTGGGGCAAGATCCGCCAAGT 1667  
1838 TCCGCGCTGCCATCCAGAAAGAGACCTCGGAGAGCTGTGGACGCTGTGGAGTACTTGGCAGGCCA 1897  
1668 TCAAGTGTCCCATCCAGAAAGAGACCTCGGAGAGCTGTGGAGCTGTGGAGTACTTGGCAGGCCA 1727  
1898 CTTGATCTCCCGAGTGGAGTTCGTGAACACCCGCCCTCTGTGAGCTGTGGTACCGAGC 1957  
1728 CTTGATCTCCCGAGTGGAGTTCGTGAACACCCGCCCTCTGTGAGCTGTGGTACCGAGC 1787  
1958 TGGAGAGGAGCCCATCTCGGCGCGGAGACCTTCTAGTGGAGCGGCGCCGCCAACCGCG 2017  
1798 TGGAGAGGAGCCCATCTGTGGCGCGGAGACCTTCTAGTGGAGCGGCGCCGCCAACCGCG 1847  
2018 AGACCAAGTCCGCAAGGCGGCTACGTGACCGGACCGGCGCGGCGGAGAGTCTGTGAGCC 2077  
1848 AGACCAAGTCCGCAAGGCGGCTACGTGACCGGACCGGCGCGGCGGAGAGTCTGTGAGCA 1907  
2078 TGACCGAGACACCAACAGAGACCGAGCTGTGAGGCCATCCAGTGTGCCCTGTGAGAGCA 2137  
1908 TGCCCGACACCAACAGAGACCGAGCTGTGAGGCCATCCAGTGTGCCCTGTGAGAGCA 1967  
2138 GCGGCGAGGAGTGAACATCTGTGACCGGACCGGACCGGCGCGGCGGAGAGTCTGTGAGCC 2197  
1968 GCGGCGTGGAGTGAACATCTGTGACCGGACCGGACCGGCGCGGAGAGTCTGTGAGCC 2027  
2198 AGCCCGACAGAGCGAGAGCGAGCTGTGTAACAGATCATCAGCAGCTGTGTAAGAGG 2257  
2028 AGCCCGACAGAGCGAGAGCGAGCTGTGTAACAGATCATCAGCAGCTGTGTAAGAGG 2087  
2258 AGAAGGTGTACCTGTGAGTGTGGTGTCCCGCCACAGAGGATCGGGGCAACAGCAGATCG 2317  
2088 AGAAGGTGTACCTGTGAGTGTGGTGTCCCGCCACAGAGGATCGGGGCAACAGCAGATCG 2147  
2318 ACAAGCTGTGAGAGGAGGATCCGCAAGTGTCTTCTGTGACGGCATCGATCGCGGCA 2377  
2148 ACAAGCTGTGAGAGGAGGATCCGCAAGTGTCTTCTGTGACGGCATCGATCGCGGCA 2207  
2378 TGTGTATCTACAGTGTACAGACGACCTGTGAGTGGGAGCGGCGGCGCTTAGGATCGATT 2437  
2208 TGTGTATCTACAGTGTACAGACGACCTGTGAGTGGGAGCGGCGGCGCTTAGGATCGATT 2267

2438 AAAAGTTCCCGGGCTAGCACCCGTTGAATTC 2469  
2268 AAAAGTTCCCGGGCTAGCACCCGTTGAATTC 2299

RESULT 7  
US-09-552-950-2  
; Sequence 2, Application US/09552950  
; Patent No. 6541248  
; GENERAL INFORMATION:  
; APPLICANT: Oxford Biomedica (UK) Limited  
; TITLE OF INVENTION: Anti-Viral Vectors  
; FILE REFERENCE: 674524-2004  
; CURRENT APPLICATION NUMBER: US/09/552,950  
; CURRENT FILING DATE: 2000-04-20  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4307  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: gagpol-SYNsp - codon  
; OTHER INFORMATION: optimised gagpol sequence  
US-09-552-950-2

Query Match 68.0%; Score 1678.6; DB 4; Length 4307;  
Best Local Similarity 82.9%; Pred. No. 5.3e-269; Mismatches 389; Indels 16; Gaps 4;  
Matches 1968; Conservative 0;

12 CATGGCCGAGGCCATGAGCCAGG---CCACAGCGCCAACTCTCTGATGACGCGCAGCAA 68  
1086 CTTGCTGAGGCCATGAGCCAGGTCAGCAATCCGCTACCATCATGATGACGCGGCGAA 1145  
69 CTTCAAGGGGCCACAGCGCATCATCAAGTCTTCACTGCGCAAGAGGGGCCACATCGC 128  
1146 CTTTCGAACCAACAGCGAAGATCTCAAGTCTTCACTGCGCAAGAGGGGCCACATCGC 1205  
129 CGCAACTGCGCGGCCCGCCCGCAAGAGGCTGTCTGGAAGTCTGCGCAAGAGGGGCCACA 188  
1206 CGCAACTGCGCGGCCCGCCCGCAAGAGGCTGTCTGGAAGTCTGCGCAAGAGGGGCCACA 1265  
189 GATGAAGGACTGACCGGCGCGCGCAGCCCACTTCTTCCGAGGAGCTGGCCTTCCCGCA 248  
1266 GATGAAGGACTGTACGAGAGAGAGCTTAA-TTTTATAGGAGAGATCTGGCCTTCTTACA 1324  
249 GGGCAAGGCGCGCGAGTTCCTCCAGGAGCAGAACCGCGCAACAGCCCCCAGCGCGCA 308  
1325 AGGAGAGCGCGAGATTTTCTTCAAGAGCAGACCGCGCAACAGCCCCCAGCGCGCA 1384  
309 GCTGCAAGT-----CGCGGCGCAACACCCCGAGCGAGGCGCGCGCGCGCGCGCGG 362  
1385 GCTTCAGGTCTGGGGTCTGGGCGCGCAACCAACTCCCGCTCCGAGCAGGAGCGCGCGG 1444  
363 CA-----CCCTGAATTCCTCCCGAGATCACCTGTGGCAGCGCCCTCTGGTGAAGCATCAA 416  
1445 CAGGTGTCTTCACTTCTCTCTCAGTCAAGTCAAGTCTTGGCAGCGACCTCTCTCAGCATCAA 1504  
417 GTTGGCGCGCGAGATCAAGAGGCGCTGTGTGACACCGCGCGCGCGCGCGCGCGCGCGG 476  
1505 GATCGGGGCGCGAGCTCAAGAGGCGCTCTCTGTGACACCGGAGCAGAGCAACCGTGTGGA 1564  
477 GGAGATGAGCTGCTCCCGCGCAAGTGGAGGCCCAAGATGATCGCGCGCATCGCGGCTTCAT 536  
1565 GGAGATGTCTGTCTCCAGGCGCTGTGAGCGCGAGATGATCGGGGAGATCGCGGCTTCAT 1624  
537 CAAGGTGCGCGAGTACAGCAGATCTGTGATCGAGATCTCGCGCAAGAGCGCATCGGCGAC 596  
1625 CAAGGTGCGCGAGTATGACAGATCTCTCATCGAAATCTCGGCGCAACAGGCTATCGGTAC 1684  
597 CTTGCTGATCGGCGG 656  
1685 CTTGCTGATCGGCGG 1744

Qy	657	CTGCACCTGAACTTCCCCATCAGCCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGG	716
Db	1745	TTGCACGCTGAACTTCCCCATTAGCCCTATCGAGACGGTACCGTGAAGCTGAAGCCCGG	1804
Qy	717	CATGGACGGCCCAAGTGTGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGCCCTGAC	776
Db	1805	GATGGACGGCCCGAAGTCAAGCATATGGCCATTTGACAGAGAGAGATCAAGGCATGGT	1864
Qy	777	CGCCATCTCGAGGAGATGAGAGAGAGGGCAAGATCACAAATCGCCCCCAGAACCC	836
Db	1865	GGAGATTGTCACAGAGATGAAAAGGAAAGGAAAAATCTCCAAGATTGGGCTCAGAACCC	1924
Qy	837	CTACAACACCCCGTGTTCGCCATCAAGAGAGAGGACAGCACCAAGTGGCGCAAGCTGGT	896
Db	1925	GTAACAACGCGCGTGTTCGCATCAAGAAAGAGGACTCGACGAAATGGCGCAAGCTGGT	1984
Qy	897	GGACTTCGCGAGCTGAACAAGGCCACCCAGGACTTCTGGAGAGTGCAGCTGGGCATCCC	956
Db	1985	GGACTTCGCGAGCTGAACAAGGCCACCGCAAGACTTCTGGAGAGTTCAGCTGGGCATCCC	2044
Qy	957	CCACCCCGCGGCTGGAAGAGAGAGAGCGTGAACCGTCTGGAGTGGCGGACGCCCTA	1016
Db	2045	GCACCCCGCAGGCTGAAGAGAGAGAAATCCGTGACCGTACTGATGTGGGTGATGCCCTA	2104
Qy	1017	CTTCAGGCTGCCCCCTGGACGAGGACTTCCGCAAGTACACGCCCTTCACCATCCCCAGCAT	1076
Db	2105	CTTCTCGGTTCCCTGGACGAAGACTTCAGAAGTACACTGCCTTCAATCCTCTCGAT	2164
Qy	1077	CAACAAAGAGACCCCGGCTCCGCTACCAAGTACAAAGTCTGCTGCCCCCAGCGCTGGAAGG	1136
Db	2165	CAACAAAGAGACACCGGGATTGCATATCAGTACAAAGTCTGCCCCCAGGCTGGAAGG	2224
Qy	1137	CAGCCCCAGGACTTTCAGAGCAGCATGACCAAGATCCTGGAGGCCCTTCGCGCCCGCAA	1196
Db	2225	CTCTCCCGCAATCTTTCAGAGTAGATGACCAAAATCCTGGAGCCTTTCGCAACAGAA	2284
Qy	1197	CCCCGAGATGTTGATCTACCAAGTACATGAGAGACCTGTACGTGGCAGCGACTGGAGAT	1256
Db	2285	CCCCGAGATGTTGATCTATCAGTACATGAGATGACTTGTACGTGGGCTCTGATCTAGAGAT	2344
Qy	1257	CGGCCAGCACCGCGCCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGGCTGGGCTTCAC	1316
Db	2345	AGGGCAGCACCGCACCAAGATCGAGGAGCTGCGCAGACACTGTTGAGTTGGGAGCTGAC	2404
Qy	1317	CACCCCGCAAGAAAGCACCAAGAGGAGCCCCCTTCTGTGATGGGCTACAGAGTGCA	1376
Db	2405	CACACCGCAAGAAAGCACCAAGAGGAGCCTCCCTTCTGTGATGGTTACGAGCTGCA	2464
Qy	1377	CCCCGCAAGTGGACCTGTCAGCCCCATCTGAGCTCCCGAGAAGAGAGCTGGACCGTGA	1436
Db	2465	CCCTGCAAAATGGACCTGACGCCCTATGTTGCTGCCAGAGAAAGACAGCTGGAATGTCAA	2524
Qy	1437	CGACATCCAGAGCTGTGGCAAGCTGAACTGGGCCAGCGCAGATCTACCCCCGCAATCAA	1496
Db	2525	CGACATACAGAACTGTGGGGAAGTTGAACTGGGCCAGTCAGATTTACCCAGGATTA	2584
Qy	1497	GGTGGCAGCTGTGCAAGCTGTGCGCGCGGCCCAAGCCCTGACCCAGACATCTGTCGCCCT	1556
Db	2585	GGTGAAGCAGCTGTGCAAACTTCTCCCGGAAACCAAGGCACTACAGAGGTGATCTCCCTCT	2644
Qy	1557	GACCGAGAGCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTTGGCGAGCCCCGTGCA	1616
Db	2645	AACCGAGAGGCCGAGCTCGAACTGGCAAGAAAACCGAGAGATCTTAAGGAGCCCGTGCA	2704
Qy	1617	CGGCGTGTACTACGACCCCAAGAGGACCTGGTGGCGCGAGATCAGAGAGCAGGSCACGA	1676
Db	2705	CGGCGTGTATATGACCCCTCAAGGACCTGATCGCCGAGATCAGAAAGCAGGGGCAAGG	2764
Qy	1677	CCAGTGGACCTTACAGATCTTACCAAGGAGCCCTTCAAGAACCTGAAGACCCGCAAGTACGC	1736
Db	2765	CCAGTGGACCTTACAGATTTTACCAAGGAGCCCTTCAAGAACCTGAAGACCCGCAAGTACGC	2824

Qy	1737	CAAGATGCGCAGCGCCACACCAACAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGAT	1799
Db	2825	CCGGATGAGGGGTGCCCACTAACACGCTCAAGCAGCTGACCGAGCGCGTGCAGAGAT	2884
Qy	1797	CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGGCTCGCCATCCAGAA	1856
Db	2885	CACCACCGAAGCATCGTGATCTGGGGAAAGACTCTCAAGTTCAAGTGCCTCCAGAA	2944
Qy	1857	GGAGACTGGAGACCTGTTGAGCCGACTACTGGCAGGCGCCTCTGATCCCCAGTGGGA	1916
Db	2945	GGAAACTGGGAACCTGGTGGACAGAGTATTGGCAGGCGCCTCTGATTTCTGAGTGGGA	3004
Qy	1917	GTTCGTGAAACACCCGCCCTCGTGTGAAGCTGTGGTACCAGCTGGAGAAGAGGCCATCAT	1976
Db	3005	GTTCGTCAACACCCCTCCCTCGTGTGAAGCTGTGGTACCAGCTGGAGAAGAGGCCATAGT	3064
Qy	1977	CGGCGCGAGACCTTTTACGTGTGACGGCGCGCCACCGCGAGACCAAGATCGCAAGGC	2036
Db	3065	GGGCGCGGAACCTTTTACGTGTGATGGGCGCGCTTAACAGGAGACTAAGCTGGGCAAGC	3124
Qy	2037	CGGCTACGTCAACCGACCGGGCGCGGAGAGATCGTGTAGCGCTCACCGAGACCCCAACCA	2096
Db	3125	CGGATAGCTCACTAAACCGGGCGAGACAGAAAGTTGTCAACCTCACTGACACCAACCA	3184
Qy	2097	GAAGCCGAGCTCGAGGCCATCCAGCTGGCGCTGCAGGACAGCGCAGCGAGTGAACAT	2156
Db	3185	GAAGACTGAGCTGGAGGCCATTTACTTCGTTTTCAGGAGCTCGGGCTGGAGGTGAACAT	3244
Qy	2157	CGTGACCGAGCGCAGTACGCCCTGGGCGCATCATCCAGGGCCAGCCCGACGAAGCGAGAG	2216
Db	3245	CGTGACAGACTCTCAGTATGCCCTCGGCGCATCATTAAGCCCGACCGACAGAGTGATC	3304
Qy	2217	CGAGCTGGTGAAACAGATCATTCAGCAGCTGTGATCAGAGGAGGAAGTGTTACCTGAGCTG	2276
Db	3305	CGAGCTGGTCAATCAGATCATCAGCAGCTGTGATCAAGAGAAAGTCTTATTTGGCCTG	3364
Qy	2277	GGTGCCCGCCCAAGAGGCGATCGCGGCGCAACGAGCAGATCGCAAGCTGGTGAAGAGG	2336
Db	3365	GGTACCCGCCCAAGAGGCAATTGGCGCAATGAGAGGTCGCAAGCTGCTCGGCTGG	3424
Qy	2337	CATCCGCAAGGTGCTGTTCTCTGACCGCATCGA	2369
Db	3425	CATCAGGAAGGTGCTATTCTCTGATGGCATCGA	3457

RESULT 8  
 US-09-552-950-5  
 ; Sequence 5, Application US/095552950  
 ; Patent No. 6541248  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Oxford Biomedica (UK) Limited  
 ; TITLE OF INVENTION: Anti-Viral Vectors  
 ; FILE REFERENCE: 674524-2004  
 ; CURRENT APPLICATION NUMBER: US/09/552,950  
 ; CURRENT FILING DATE: 2000-04-20  
 ; NUMBER OF SEQ ID NOS: 22  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 9772  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP  
 ; US-09-552-950-5

Query Match	66.9%	Score 1651.4	DB 4	Length 9772
Best Local Similarity	82.2%	Pred. No. 1.8e-264		
Matches 1951	Conservative	0	Mismatches 406	Indels 16
Gaps	4			
QY	12	CATGGCGAGCCATGATGACCCAGG---CCACGAGGCCCAATCTGTGATGAGCGCGACGAA	68	
DP	2193	CCTGGCTGAGGCCATGATGAGCGAGGTGACCACTCCGCTACCATCATGATGAGCGCGCGCAA	2252	

QY 69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGGCGCAAGGAGGCGCATCGC 128  
D5 2253 CTTTCGGAACCAACGCAAGATCGTCAAGTGTCTTCAACTGGCGCAAGGAGGCGCATCGC 2312  
QY 129 CGCAACTCGCCGCGCCCGCCGCAAGAGGCTGTCTGGAAGTGCAGCAAGGAGGCGCCACA 188  
D5 2313 CGCAACTCGAGGGCCCTAGGAAAGAGGCTGTGGAAGTGTGGAAGGAGGAGGAGCACA 2372  
QY 189 GATGAAGAGTCTACCGAGCGCCAGGCGCAATCTCTTCGCGAGGACTTGGGCTTCCCGCA 248  
D5 2373 AATGAAGATTGTACTGAGAGCAGGCTAA-TTTTITAGGGAAGATCTGGGCTTCCACA 2431  
QY 249 GGGCAAGGCCCGCGAGTTCCTCCAGCGAGCAGAAACCGCGCAACGCCCAACAGCGCGGA 308  
D5 2432 AGGGAAGGCCAGGGAATTTCTTACAGACAGACAGACAGCCCAACAGGAGAGA 2491  
QY 309 GGTGAGAGTGCCTGCGG-----CGACAAACCCCGCAGCGAGGCGCGCGAGCGCAGGG 362  
D5 2492 GCTTCAGGTTTGGGAAGAGACAACAACCTCCCTCTCAGAAGCAGGAGCGGATAGCAAGG 2551  
QY 363 CA-----CCCTGAATTCCTCCAGATCACTCTTGGCAGGACCCCTCTGTCACATANA 2611  
D5 2552 AACTGTATCTTTAGCTTCCCTCAGATCACTCTTGGCAGGACCCCTCTGTCACATANA 2611  
QY 417 GGTGGCGGCGAGATCAAGAGGCGCTCTGTGACACCGCGCGCAGACACCGTGTGGA 476  
D5 2612 GATAGGGGGGAGCTCAAGAGGCTCTCTGTGACACCGGAGCAGACAGCACCGTGTGGA 2671  
QY 477 GAGATGAGCTGCGCGCAAGTGAAGCGCAAGATGATCGCGCATCGCGGCTTCAT 536  
D5 2672 GAGATGCTGTCACAGCGCTGGAAGCGGAGATGATCGGGGAATCGCGGTTTCAT 2731  
QY 537 CAAGTGGCGCAGTACGACAGATCCTGATCGAGATCTGCGCAAGAGGCGCATCGGCAC 596  
D5 2732 CAAGTGGCGCAGTATGACCAATCCTCATCGAAATCTGCGGCCACAAGGCTATCGGTAC 2791  
QY 597 GGTGCTGATCGCGCCGACCCCGTGAAATCATCGCGCGCAGACATGTCAGCCAGCTGG 656  
D5 2792 GGTGCTGATCGCGCCGACACCGCTCAACATCATCGGACGCAACCTGTGACGCGATCG 2851  
QY 657 CTGCACTCTGAACTTCCCATCAGCCCATCGAGACCGTGGCGTGAAGCTGAAGCCCGG 716  
D5 2852 TTGACGCTGAACTTCCCATTAGCCCTATCGAGACGCTACCGGTGAAGCTGAAGCCCGG 2911  
QY 717 GATGACCGCCCGCAAGTGAAGTGGCCCTGACCGAGAGAGATCAAGGCGCTGAC 776  
D5 2912 GATGACCGCCCGCAAGTGAAGTGGCCCTGACCGAGAGAGATCAAGGCGCTGAC 2971  
QY 777 CGCCATCTGCGAGGAGTGGAGAGGCGCAAGATCACCAAGATCGGCCCGGAGACCC 836  
D5 2972 GAGATTTGCAAGAGTGGAAAGAGAGGAGAAATCTCAGATTTGGGCTGAGAACCC 3031  
QY 837 CTACACACCCCGCTTTCGCGCATCAAGAGAGAGACAGCAACCAAGTGGCGCAAGCTGT 896  
D5 3032 GTACACACCGCGGTGTTCCGAATCAAGAGAGGACTCGACGAATGGCGCAAGCTGT 3091  
QY 897 GGACTTCGCGAGCTGAACAGCGCACCCAGACTTCTGGAGGTGAGCTGGGCAATCCC 956  
D5 3092 GGACTTCGCGAGCTGAACAGCGCACCCAGACTTCTGGAGGTGAGCTGGGCAATCCC 3151  
QY 957 CCACCCCGCGCTTGAAGAGAGAGAGCGTGAACCGTGTGGAAGTGGCGGAGCGCTA 1016  
D5 3152 GCACCCCGCAGGCTGAAGAGAGAGAAATCCGTGACCGCTACTGAGTGTGGGTGAGCTA 3211  
QY 1017 CTTAGCGTGGCTGGAGGAGGACTTCGCGAGTACACCGGCTTCAACATCCCGAGAT 1076  
D5 3212 CTTTCGCTTCCCTTGGAGAGAGATTGAGGAGTACATCGCTTCAAAATCCCTTCGAT 3271  
QY 1077 CAACACAGAGACCCCGGCTCCGCTACAGTACAACTGTGTCGCCAGGCGCTGAAGGG 1136  
D5 3272 CAACACAGAGACCCCGGCTTTCGATATCAGTACAACTGTGTCGCCAGGCGCTGAAGGG 3331  
QY 1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCGAGACCTTCCCGCGCCCGCA 1196

D5 3332 CTTCCCGCAATCTTCCAGAGTAGATGACCAAAATCTCGAGACCTTTCGCGCAACAGAA 3391  
QY 1197 CCCGAGATCGTGTATCTACAGTACATGACACCTGTACGTGGGAGCGACTCGAGAT 1256  
D5 3392 CCCGACATCGTCTATCTAGTACATGATGACTGTACGTGGGCTCTGATCTAGAGAT 3451  
QY 1257 CGSCAGCACCGCGCAAGATCGAGAGCTGGCAAGCACTCTGCTGCGCTGGGCTTCAC 1316  
D5 3452 AGGCGAGCACCGCACCAAGATCGAGAGCTGGCCAGCACTGTGTAGGTGGGACTGAC 3511  
QY 1317 CACCCCGCAAGAACACACAGAGAGCCCTTCTCTGTGTGGATGGGCTACGAGTGA 1376  
D5 3512 CACACCGCAAGAACACACAGAGAGCCCTCTCTCTCTGATGGGTTCAGAGTGA 3571  
QY 1377 CCCGCAAGTGGACCGTGTAGCCCATCGAGCTGCCGAGAGAGAGTGAACGTGA 1436  
D5 3572 CCTGCAAAATGACCGTGTAGCCCTATCTGTGTCCAGAGAAAGACAGCTGACTGTCAA 3631  
QY 1437 CGACATCCAGAACTGCTGGGCAAGCTGAACTGGGCGACCCAGATCTACCCCGGATCAA 1496  
D5 3632 CGACATACAGAGCTGCTGGGAGTTGAACTGGGCCAGTCAAGTTTACCAGGATTA 3691  
QY 1497 GGTGCGCAGCTGTGCAAGCTGCTGGCGCGCAAGGCCCTTGACCGACATCGTGCCT 1556  
D5 3692 GGTGAGGCGAGCTGTGCAAACTCTCTCGCGGAAACCAAGGCACCTCACAGAGGTGATCCCT 3751  
QY 1557 GACCGAGGAGCGCGAGCTGAGCTGCGCGAGAAACCGCGAGATCTCTGCGGAGCCCGTGA 1616  
D5 3752 AACCGAGAGGCGCGAGCTGAACTGCGCAAGAAACCGAGAGATCTTAAGAGAGCCCGTGA 3811  
QY 1617 CGCGCTGTACTAGCACCCAGCAAGACCTGTGGCGCAGATCCAGAGCGGCGCAACGA 1676  
D5 3812 CGCGCTGTACTATGACCCCTTCAAGGACCTGATCGCGAGATCCAGAGCGGCGCAAG 3871  
QY 1677 CAGTGGACCTTACAGATCTACAGAGGCGCTTCAAGACCTGAGACCGGCAAGTACG 1736  
D5 3872 CAGTGGACCTTACAGATTTACAGAGGCGCTTCAAGAACCTTGAAGACCGGCAAGTACG 3931  
QY 1737 CAAGATGCGCACCGCGCCACACCAACGACGTGAAGCAGCTGACCGAGGCGCTGCAGAGAT 1796  
D5 3932 CCGGATGAGGGGTGCCCACTAAACGACGTGAAGCAGCTGACCGAGGCGCTGCAGAGAT 3991  
QY 1797 CGCATGAGAGAGATCTGTATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAA 1856  
D5 3992 CACACCGCAAGAGATCTGTATCTGGGGAAAGACTCTTAAGTTTCAAGCTGCCCATCCAGAA 4051  
QY 1857 GAGAGCTGGAGACCTGTGTGACCCGACTACTGGGAGCGCACCTGATCCCGAGTGGGA 1916  
D5 4052 GGAACCTGGGAAACCTGTGTGACAGATTTGGCAGGCCACTGTGATTCCTGAGTGGGA 4111  
QY 1917 GTTCGTGAACACCCCGCCCTGTGTGAAGCTGTGTGTAACAGCTGAGAGAGAGCCCATCAT 1976  
D5 4112 GTTCGTGAACACCCCGCCCTGTGTGAAGCTGTGTGTAACAGCTGAGAGAGAGCCCATCAT 4171  
QY 1977 CGSCCGCGAGACCTTCTAGTGAAGCGCGCCGCAACCGCGAGACCAAGATCGGCAAGCG 2036  
D5 4172 GGGCGCGGAAACCTTCTAGTGTGATGGGCGCGCTAACAGGAGAGATTAAGCTGGGCAAGC 4231  
QY 2037 CGGCTACGTGAACGACCGCGGCGCGAGAGATCTGTAGCTGTGACCGAGACCAACCAACA 2096  
D5 4232 CGGATACGTCACTAACCGGCGGAGACAGAAAGGTGTGACCCCTCACTGACACCAACCAACA 4291  
QY 2097 GAGAGCGGAGCTGACGCGCATCCAGCTGGCCCTCGAGAGACAGCGGCGAGGCTGAACAT 2156  
D5 4292 GAGAGCTGAGCTGACGCGCATTTTACTCTGAGGAGCTCGGCGCTGAGAGTGAACAT 4351  
QY 2157 CGTACCGCAGCAGCAGTACGCGCATCTCCAGGCGCCAGCCCGCAAGAGCGAGAG 2216  
D5 4352 CGTGACAGACTCTCAGTATGCGCTGGCATCTTAAGCCCGCAGACCAAGTGAAGTC 4411  
QY 2217 CGAGCTGTGTGAACAGATCATCGAGCAGCTGATCAAGAGAGAGAGGTGATCTGAGCTG 2276

```

4412 CGAGCTGGTCAATCAGATCATCGACAGCTGATCAAGAGGAAAAGGTATCTGCCTG 4471
2277 GGTGCCCGCCCAACAGGGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTGAGCAAGG 2336
4472 GGTACCGCCCAACAGGCATTCGGCGCATGAGCAGGTCGACAAGCTGGTCTCGCGTGG 4531
2337 CATCCGACAGGTGCTGTTCTCTGACGCGCATCGA 2369
4532 CATCAGAGAGGTGCTATTCCTGGATGGCATCGA 4564

RESULT 9
US-09-872-733A-6
; Sequence 6, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL,
; TITLE OF INVENTION: HIV ENV GENES
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL, HIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733A
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 8366
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA sequence
; OTHER INFORMATION: of the construct pCMVgagpolNkan containing a CMV
; OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin
; OTHER INFORMATION: resistance gene
US-09-872-733A-6

```

2276	TCGGGGGGAAC	TCAAGNAGCGCTGCTCGATACAGGACGACATGATACAGTATTAGAAG	2333
479	AGATGAGCCTGCCCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGGCTTCATCA	538	
2336	AAATGATTTGTCAGGAAGATGGAAACCAAAAATGATAGCGGGGATTCGGGGCTTCATCA	2395	
539	AGGTGGCCAGTACGACACAGATCCCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCG	598	
2396	AGGTGGCGAGTACGACCGAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAG	2455	
599	TGCTGATCGGCCCCACCCCGTGAAACAATCATCGGCGGCAACATGCTGTGACCCAGCTGGGCT	658	
2456	TATTAGTAGGACCTACACCTGTCAACATAAATTGGAAAGAAATCTGTTGACCCAGATCGGCT	2515	
659	GCACCTTGAACCTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGGAAGCCCGGCA	718	
2516	GCACCTTGAACCTTCCCATCAGCCCTATTGAGACGGTCCCGTGAAAGTTGAAGCCCGGGA	2575	
719	TGACGCCCCCAAGGTGAACGATGGCCCTCAGCCAGAGAGAGATCAAGGCCCTGACCG	778	
2576	TGGACGGCCCAAGGTCAGCAATGGCCATTGACGAAGAAGAGATCAAGGCTTAGTGG	2635	
779	CNATCTGGAGGAGATGGAGAGAGGGGCAAGATCACCAAGATCGGCCCGGAGAACCCCT	838	
2636	AAATCTGTACAGAGATGGAGAAGGAAGATCAGCAAGATCGGGCTTGAAACCCCT	2695	
839	ACAACACCCCGTGTTCGCCATCAAGAAAGACACGACCAAGTGGCGCAAGCTGGTG	898	
2696	ACAAACACTCCAGTCTCGCAATCAAGAAGAAGACAGTACCAAGTGGAGAAAGCTGGTG	2755	
899	ACTTCCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCATCCCC	958	
2756	ACTTCAGAGAGCTGAACAAGAGAACTCAGAGACTTCTGGGAAGTTACGTGGGCATCCCC	2815	
959	ACCCCGCCGCTGGAAGAGAGAGAGAGCGTGACCGTGTGGAAGTGGCGAGCGCCTACT	1018	
2816	ATCCCGCTGGTGTGAAGAAGAAGATCAGTGACGTGTGATGTGGGTGATGCCCTACT	2875	
1019	TCAGCTGCCCCTGGACGAGGACTTCCGAAGTACACCGCTTACCATCCCAGCATCA	1078	
2876	TCTCCGTTCCCTTGGACGAGGACTTCAGGAAGTACACTGCCCTCAGGATACCTAGCATCA	2935	
1079	ACAAACAGACCCCGCGCATCCGTACACGATCAAAAGTGTGCCCCAGAGGTGGAAGGGCA	1138	
2936	ACAACAGACACCAAGCATCCCGTACACGATCAACGTGCTGCCACAGGATGGAAGGGAT	2995	
1139	GCCCCAGCATCTTCCAGAGCAGCATACCAAGATCTGGAGCCCTTCCGCGCCCGCAACC	1198	
2996	CACAGCCATCTTTCAAAGCAGCATGACCAAGATCTGGAGCCCTTCGCAAGCAAAACC	3055	
1199	CCGAGATCGTGATCTTACAGTACATGGAAGACTGTGAGTGGCGAGGACCTGGAGATCG	1258	
3056	CAGACATCTGTGATCTTACGTATCATGGACACCTCTACGTAGGAAGTGAACCTGGAGATCG	3115	
1259	GCACAGCACGCGCAAGATCGAGGAGTGGCAAGCACTGCTGCGCTGGGGCTTCACCA	1318	
3116	GGCAGCACAGGACCAAGATCGAGGAGTGAAGCAGCATCTGTTGAGTGGGACTGACCA	3175	
1319	CCCCGACAAAGACCAAGAGAGGCCCTTCTGTGGATGGGGTACGAGCTGGACC	1378	
3176	CACACAGAAAGCACCAAGAGAACTCCCTTCTGTGGATGGGCTACGAACTGCATC	3235	
1379	CCGACAGTGGACGCTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGACACCGTGAACG	1438	
3236	CTGACAAGTGGACAGTGCAGCCCATGTGCTGCTGAGAAAGGACAGCTGTGNAACG	3295	
1439	ACATCCAGAAAGCTGTGGGCAAGCTGAATGGGCGACCGCAGATCTACCCCGGCATCAAGG	1498	
3296	ACATACAGAGCTCGTGGGCAAGTTGAATGGGCAAGCCAGATCTACCAAGGATCAAG	3355	
1499	TGGCGACAGCTGTGCAAGCTCTCGCGCGCGCAAGGCCCTGACCGCATCGTGCCCTCGA	1558	
3356	TTAGGCAGCTGTGCAAGCTCTTCGAGGAACCAAGGCCACTGACAGAAAGTGAATCCACTGA	3415	

1559 CCGAGGAGCGAGCTGGAGTCCCGAGACCGGAGATCCCTCGCGAGCCCGTGCACG 1618  
1560 CCGAGGAGCGAGCTGGAGTCCCGAGACCGGAGATCCCTCGCGAGCCCGTGCACG 1619  
3416 CAGAGGAGCGAGCTAGAACTGGCAGAGACCGGAGATCCCTGAAGAGCCAGTACATG 3475  
3417 CAGAGGAGCGAGCTAGAACTGGCAGAGACCGGAGATCCCTGAAGAGCCAGTACATG 3476  
1619 CCGTGTACTACGACCCCGAGCAAGGACCTGGTGGCCGAGATCCAGAGCAGGCGCACGACC 1678  
1620 CCGTGTACTACGACCCCGAGCAAGGACCTGGTGGCCGAGATCCAGAGCAGGCGCACGACC 1679  
3476 GAGTGTACTACGACCCCGAGCAAGGACCTGGTGGCCGAGATCCAGAGCAGGCGCACGACC 3535  
1679 AGTGACCTACAGATCTACAGAGGCGCTTCAAGAACCTGGAACCGGCAAGTACGCCA 1738  
1680 AGTGACCTACAGATCTACAGAGGCGCTTCAAGAACCTGGAACCGGCAAGTACGCCA 1739  
3536 AATGGACCTACCAAACTTACAGAGGCGCTTCAAGAACCTGGAACCGGCAAGTACGCCA 3595  
1739 AGATGCGCACCGCCACACCAAGCAGTGAAGCAGCTGACCGAGGCGCTGAGAGGATCG 1798  
3596 GATGAGGGTGGCCACACCAAGCAGTGAAGCAGCTGACCGAGGCGCTGAGAGGATCG 3655  
1799 CCATGGAGAGATCTGTGATCTGGGGCAAGACCCCGAGTTCGGCTGCGCCATCCAGAGG 1858  
1800 CCATGGAGAGATCTGTGATCTGGGGCAAGACCCCGAGTTCGGCTGCGCCATCCAGAGG 1859  
3656 CCACAGAGAGATCTGTGATCTGGGGCAAGACCCCGAGTTCGGCTGCGCCATCCAGAGG 3715  
1859 AGACCTGGAGAGCTGTGGGCAAGCAGTGAAGCAGCTGACCGAGGCGCTGAGAGGATCG 1918  
3716 AGACATGGAGAGATCTGTGATCTGGGGCAAGACCCCGAGTTCGGCTGCGCCATCCAGAGG 3775  
1919 TCGTGAACACCCCGCTGGTGAAGCTGTGTACCAAGCTGGAAGAGGCGCCATCCATCG 1978  
3776 TCGTGAACACCCCGCTGGTGAAGCTGTGTACCAAGCTGGAAGAGGCGCCATCCATCG 3835  
1979 GCGCGAGAGCTTCTACCTGGAGCGCGCGCCACCGGAGACCAAGATCGGCAAGGCGCG 2038  
3836 GAGCAGAGACCTTCTACCTGGAGCGCGCGCCACCGGAGACCAAGATCGGCAAGGCGCG 3895  
2039 GCTACGTGACCGAGCGCGCGCGCGCCACCGGAGACCAAGATCGGCAAGGCGCG 2098  
3896 GCTACGTGACCGAGCGCGCGCGCGCCACCGGAGACCAAGATCGGCAAGGCGCG 3955  
2099 AGACCGAGCTGAGCGCGCTACAGTGGCGCTGCGAGGAGCGCGAGCTGAGTACCTGGG 2158  
3956 AGACTGTGAAGCGCGCTACCTACCTGCTGCTGAGAGCGCGAGCTGAGTACCTGGG 4015  
2159 TGACCGAGCGCGCGCTGCGCGCTGCGAGCTACCTGCGCGCGCGCGCGCGCGCGCGCG 2218  
4016 TGACGAGCTCAAGTACCGCGCTGCGCGCTGCGAGCGCGCGCGCGCGCGCGCGCGCG 4075  
2219 AGCTGTGAGCGAGCTACGAGCGAGCTGATCAAGAGGAGAGTGTACCTGAGCTGGG 2278  
4076 AGCTGTGAGCGAGCTACGAGCGAGCTGATCAAGAGGAGAGTGTACCTGAGCTGGG 4135  
2279 TGCG 2338  
4136 TACGAGCAGCAGAGGAGTGGAGGAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAG 4195  
2339 TCGCAAGGTGTGTCTTCTGGAGCGGATCGAT 2370  
4196 TCGGAGAGGTGTGTCTTCTGGAGCGGATCGAT 4227

RESULT 10  
US-09-872-733A-1  
; Sequence 1, Application US/09872733A  
; Patent No. 6656706  
; GENERAL INFORMATION:  
; APPLICANT: The Government of the United States of America, as  
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, HIV GAG AND  
; TITLE OF INVENTION: HIV ENV GENES  
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL, HIV GAG & ENV  
; CURRENT APPLICATION NUMBER: US/09/872,733A  
; PRIOR FILING DATE: 2001-06-01  
; PRIOR APPLICATION NUMBER: PCT/US00/34985  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/173,036  
; PRIOR FILING DATE: 1999-12-23

; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4338  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Mutated Human  
; IMMUNODEFICIENCY VIRUS - 1 Gag/Pol gene  
US-09-872-733A-1  
Query Match 63.1%; Score 1557.4; DB 4; Length 4338;  
Best Local Similarity 80.5%; Pred. No. 5.4e-249;  
Matches 1914; Conservative 0; Mismatches 441; Indels 22; Gaps 7;  
QY 14 TGGCCGAGCGCCATGAGCCAGGCGCACAGC---GCCAACATCTGTATGAGCGCAGCAACT 70  
DB 1085 TGGCCGAGCGCGATGAGCCAGGCGTACGAACTCGCGGACCATATGATGAGAGGCAACT 1144  
QY 71 TCAGGGGCCCCAAGCGCATCATCAAGTGTCTCAAGTGTCTGCGGCAAGAGGCGCCACATCGCCC 130  
DB 1145 TCGGAACACAGCGGGAAGATCGTCAAGTGTCTCAATTTGGGCAAGAGGCGCACCGCCA 1204  
QY 131 GCAACTGCG 190  
DB 1205 GGNACTGCG 1264  
QY 191 TGAAGACTGTACCGAGCG 250  
DB 1265 TGAAGATTTGATGTAGAGACAGGCTAA-TTTTGTAGGGAAGATCTGGGCTTTCTTACAG 1323  
QY 251 GCAAGCG 310  
DB 1324 GGAAGCG 1383  
QY 311 TGCAGTGTGCGCGG-----CGACAAACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 364  
DB 1384 TGCAGTGTGCGCGGCG 1443  
QY 365 -----CCCTGACCTTCCCGCGAGTCAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCG 418  
DB 1444 CTGTATCTCTTTAACTTCT 1503  
QY 419 TGGCGCGCGCGAGTCAAGAGGCGCGCTCTCTGACACCGCGCGCGCGCGCGCGCGCGCGCG 478  
DB 1504 TCGGGGCG 1563  
QY 479 AGATGAGCTGTCCCG 538  
DB 1564 AAATGAGTTTTCGCGGAGAGTGGAAACCAAAATGATGCGGGGGGATCGGGGGGCTTCA 1623  
QY 539 AGGTGCGCGCGAGTACGACCGAGTCTCTGATCGAGATCTGCGCGCGCGCGCGCGCGCGCG 598  
DB 1624 AGGTGAGCGAGTACGACCGAGTCTCTGATAGTATCTGTGACATAGCTATAGGTACAG 1683  
QY 599 TGTGTATCG 654  
DB 1684 TATTAGTAGGACCTACCTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1743  
QY 655 GGTGTGACCGCTGAACTTCCCGCGAGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 714  
DB 1744 GGTGTGACCGCTGAACTTCCCGCGAGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1803  
QY 715 GGCATGACCG 774  
DB 1804 GGCATGACCG 1863  
QY 775 ACCGCGATCTGCGAGGAGTGGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 834  
DB 1864 GTCGAAATCTGTACAGAGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1923  
QY 835 CCCTTACCAACCG 894



1974 CATCGGCGCGGACCTTCTACGTGGACGGCGCGCCAAACCGGAGACCAAGATCGGCAA 2033  
3064 CGTGGGAGCAGAGACCTTCTACGTGGATGGGCGAGCCACAGGAGACCAAGCTGGGCAA 3123  
2034 GCGCGGCTTACGTAGCCAGCGGGCGCGGAGAGATCGTGAGCTGACCGGACCAACCAA 2093  
3124 GCGAGGCTTACGTAGCCAAACCGAGGACGACAGAAAGTGGTACCTGACTGACACCAACAA 3183  
2094 CCAGAGACCGAGCTGACGAGCCATCCAGCTGGCCCTGAGGACGAGCGGACGAGGTGAA 2153  
3184 CCAGAGACTGAGTGAAGCCATCTACCTAGCTTCTGAAAGACGCGGATGGAGTGA 3243  
2154 CATGCTGACCCAGCAGCAGTACGCTGGGATCATCCAGGCCAGCCCGGACCAAGAGCGA 2213  
3244 CATGCTGACAGACTCACAGTAGC-CATGGGCATCATCCAAAGCACACAGACCAATCCGA 3302  
2214 GAGCGAGCTGTAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAG 2273  
3303 GTCAGAGCTGTAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAG 3362  
2274 CTGGGTGCGCCCGCCAGGAGGATCGGCGGCAACGAGCAGATCGAAGCTGTGTGAGCAA 2333  
3363 ATGGGTACAGCAGACACAGGAAATGGAGGAAATGAACAGTAGATAAATTAGTCAGTGC 3422  
2334 GGGCATCCGCAAGGTGCTGTTCTTGGAGCGCATCGAT 2370  
3423 TGGGATCCGGAAGGTGCTGTTCTTGGAGCGGATCGAT 3459

RESULT 11  
US-09-184-418C-8  
; Sequence 8, Application US/09184418C  
; Patent No. 6492110  
; GENERAL INFORMATION:  
; APPLICANT: Hahn, Beatrice  
; APPLICANT: Gao, Feng  
; APPLICANT: Shaw, George  
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN  
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1  
; FILE REFERENCE: D6287  
; CURRENT APPLICATION NUMBER: US/09/184,418C  
; CURRENT FILING DATE: 1999-11-02  
; NUMBER OF SEQ ID NOS: 112  
; SEQ ID NO 8  
; LENGTH: 9010  
; TYPE: DNA  
; ORGANISM: Human immunodeficiency virus type 1  
; FEATURE:  
; OTHER INFORMATION: isolate=96ZM651; 137..1621:"gag"; 1426..4425:"pol";  
; OTHER INFORMATION: 4370..4948:"vif"; 4888..5178:"vpr";  
; OTHER INFORMATION: 5159..5373-7734..7824:"tat"; 5298..5373-7734..7981:"rev";  
; OTHER INFORMATION: 5387..5647:"vpu"; 5565..8171:"env"; 8173..8793:"nef"  
US-09-184-418C-8

Query Match 49.7%; Score 1227.2; DB 4; Length 9010;  
Best Local Similarity 70.3%; Pred. No. 2e-194;  
Matches 1657; Conservative 0; Mismatches 698; Indels 1; Gaps 1;

14 TGCGCGAGGCCATGAGCGGCGCCAGCGCCAGCATCTGATGCGCGGAGCAACTTCA 73  
1221 TGCTGAGCAATGAGCGCAACAAATAGTGAACATGATGCGAAGAGCAATTTA 1280  
74 AGGGCCCCCAAGCGCATCATAGTCTTCAACTGCGGCAAGAGGGGCCACATCGCCCCCA 133  
1281 AAGGAAATAAAAGAAATGGTTAAATGTTTAACTGTTGTAAGGAGGCGACATAGCGAA 1340  
134 ACTGCGCGCCCCCGCGAGAGGGCTGCTGAACTGCGGCAAGGAGGCGCCACAGATGA 193  
1341 ATTGAGGGCCCCCTAGGAAAAAGGGCTGTTGGAATGTGGAAGGAGGAGGACCAATGA 1400  
194 AGGACTGCAACCGAGCGCGGCGCAACTTCTTCCGCGAGCACTGCGCTTCCCGGCGCA 253

1924 CCCTAACAACTCCAGTCTTCGCAATCAAGAGAGGACAGTACCAAGTGGAGAAAGCTG 1983  
895 GTGACTTCCGCGAGCTGAAACAAGCGCACCCAGGACTTCTTGGAGAGTGCAGCTGGGCATC 954  
1984 GTGACTTTCAGAGAGCTGAAACAAGAGAACTCAGGACTTCTTGGGAAGTTCAGCTGGGCATC 2043  
955 CCCCACCCCGCGGCTTGAAGAGAGAGAGAGCGTGAACCGTGTGAGAGTGGCGGACGCG 1014  
2044 CCACATCCCGCTGGTGTGAAGAGAGAGAACTCAAGTGCAGTGTGAGTGTGGTGTATGCC 2103  
1015 TACTTACGCTGCGCTTGGAGAGAGACTTCCGCAAGTACACCGCTTACCATCCCGAGC 1074  
2104 TACTTCCCGTTCCTTGGAGAGAGACTTCCAGGAAGTACACTGCCCTTACGATACCTAGC 2163  
1075 ATCAACAGAGAGACCCCGGCTACCGTACAGTACAAAGTGTGCGCCAGAGGCTGGAAG 1134  
2164 ATCAACAGAGAGACCCAGGATCCGCTACAGTACAAAGTGTGCGCAAGGAGTGAAG 2223  
1135 GGCAGCCCGACATCTTCCAGAGCAGCATCAACAGATCTGAGAGCCCTTCCCGCGCCGC 1194  
2224 GGATCACAGCCATCTTCAAGCAGCATCAACAGATCTGAGAGCCCTTCCGCAAGCAA 2283  
1195 AACCCTGAGTCTGATCTACAGTACAGGAGCTGATGAGGAGGAGGAGGAGGAG 1254  
2284 AACCAGACATCTGATCTATCAGTACATGAGCAGCTTACGTAGGAAGTGAAGTGAAG 2343  
1255 ATC-GGCCAGCACCGCGCCAGAGTTCGAGAGGAGTTCGCGCAAGCACTGCTGCGCTGGGCTT 1313  
2344 ATCGGGGAGCACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGACT 2403  
1314 CACCACCCCGAGAGAGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1373  
2404 GACCACACAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2463  
1374 GCACCCCGACAAAGTGGACCGTGCAGGCGCATCGAGCTGCCGAGAGGAGGAGGAGGAGGAG 1433  
2464 GCATCCTGACAAAGTGGAGAGTGCAGGCGCATCTGCTGCTGAGAGGAGGAGGAGGAGGAG 2523  
1434 GAACGATCAGAGAGTGTGGGAGAGTGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1493  
2524 GAACGATCAGAGAGTGTGGGAGAGTGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2583  
1494 CAAGTGGCGGAGTGTGAGAGTGTGCGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1553  
2584 CAAGTGGAGAGTGTGAGAGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2643  
1554 CTTGACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1613  
2644 ACTGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2703  
1614 GCAGGCTGTACTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1673  
2704 ACATGGAGTGTACTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2763  
1674 CGACAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1733  
2764 AGGCAATGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2823  
1734 CGCAAGATGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1793  
2824 CGCAAGATGAGGGGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2883  
1794 GATCGGCAATGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1853  
2884 GATCACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2943  
1854 GAAGGAG 1913  
2944 GAAGGAG 3003  
1914 GGAGTTCGTGAACACCCCGGCTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1973  
3004 GGAGTTCGTGAACACCCCGGCTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3063

Db 1401 AAGACTGTACTGAGAGGAGGCTAA-TTTTTTAGGAAAAATTTGGCCTTCCCAAGGGA 1459  
Qy 254 AGGCCCGAGTCCCGAGCGAGCAGAACCGCGCCAAACAGCCACCAGCGCGAGCTGC 313  
Db 1460 AGGCCAGGAAATTTCTTCAGAACAGGCCAGAGCCAAAGCCGCCAGCAGAGAGCTTC 1519  
Qy 314 AGGTGCGGGGGA CAACCCCGCAGAGGAGCGCGCGCGAGCGCGAGGCAACCTGAAT 373  
Db 1520 AGGTTGAGGAGACAACCCCGCTCCGAAGCAGAGTCCGAAGACAGGGAAGCTTAACT 1579  
Qy 374 TCCCCCAGATCACCTGTGGCAGCGCCCTCTGGTGAGCATCAAGGTGGCGGCCAGATCA 433  
Db 1580 TCCCTCAATTCATCTTTGGCAGCGAACCCCTTGCTCAATAAGGTAGGGGCGAATAA 1639  
Qy 434 AGGAGGCTCTCTGGACACCGCGCCCGACGACACCGCTGTGGAGGAGATGAGCTGCCCG 493  
Db 1640 AGGAGGCTCTCTTAGACACCGGAGCAGGTGATACAGTATTAGAGAAATAAATTTGCCAG 1699  
Qy 494 GCAAGTGGAGCCCAAGATGATCGCGGCATCGCGGCTTCATCAAGGTGGCGCAGTAG 553  
Db 1700 GCAATGGAAACCAAAATGATAGGAGAAATGGAGGCTTTATCGAAGTAAGCAATATG 1759  
Qy 554 ACCAGATCTCTGATCGAGATCTCGCGCAAGAGGCCATCGGCACCGTCTGATCGGCCCA 613  
Db 1760 ATCAATACCTATGGAATTTGTGGAATAAGGCTATAGGTACAGTATTAGTAGGACCTA 1819  
Qy 614 CCCCCTGAAATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACTGAACTTCC 673  
Db 1820 CACCTGTCAACATAAATTTGGAGAAATATGTTGACTCAGCTTGGATGACACATAAATTTTC 1879  
Qy 674 CCATCAGCGCCCATCGAGACCGTCTGAGCTGAGCGCGGCGGATCGAGCGGCCCAAGG 733  
Db 1880 CAATTAGTCTTATGAACTGTACAGTAAATTTAAAGCCAGGATGATGGTGGCCCAAGG 1939  
Qy 734 TGAAGCAGTGGCCCTGACCGAGAGAGATCAAGCGCTGACCGCCATCTGCGAGGAGA 793  
Db 1940 TTAACCAATGGCCATTGACAGAGAGAAATAAAGCTTTAAACAGCAATTTGTGAAGAA 1999  
Qy 794 TGGAGAGGAGGCGCAGATCACCAAGATCGCGCCCGAGAACCCCTCAACACCCCGGTGT 853  
Db 2000 TGGAGAGGAGGAGAAATTTACAAATTTGGGCTGAAATTCATATAACACTCCAGTAT 2059  
Qy 854 TCGCCATCAAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGGATTTCCGGGAGCTGA 913  
Db 2060 TTGCCATAAAGAGAGAGCAGTACTAAGTGGCGAAATTTAGTAGATTTTCAGGGAATCA 2119  
Qy 914 ACAGCGCCAGCAGATCTTGGAGGTGCACTGGGCGATCCCCACCGCGCGCGCTGA 973  
Db 2120 ATAAAGAACTCAAGACTTTTGGAGGTTCAATTAGGAATACCAACCCAGCGAGGTTAA 2179  
Qy 974 AGAAGAGAGAGCGGTGACCGTGTGGACGTGGGCGACGCTACTTCAGCGTGGCCCTGG 1033  
Db 2180 AAAAGAAAAATCAGTGACAGTACTGGATGTGGGGGATGCATATTTTCAGTTCCCTTAG 2239  
Qy 1034 ACGAGGACTTCGCAAGTACACCGCTTACCATCCCGCAGCATCAACAGAGACCCCG 1093  
Db 2240 ATGAAAGCTTCAGGAAATATATCTCATTTCCATACCTAGTACAAACAATGAAACACCA 2299  
Qy 1094 GCATCCGCTACAGTACAAAGTGTGCGCCAGGCTGGAAGGCGAGCCCGCAGATCTTCC 1153  
Db 2300 GGATTAGATCAATATAATGTGCTTCCACAGGATGGAAGGATCACAGCAATATTC 2359  
Qy 1154 AGACAGCATGACCAAGATCTGAGCGCTTTCGCGCCCGCAACCCCGAGATCGTATCT 1213  
Db 2360 AGATAGCATGACAAAAATCTTAGAGCCCTTCAGGGCACAATAATCCAGACATAGTCTCT 2419  
Qy 1214 ACCAGTACATGAGCAGCTGTACGTGGGCGAGCTGAGAGTCGGCGCAGCACCGGCCA 1273  
Db 2420 ATCAATATATGGATGACCTGTATGTAGATCTGACTTAGAATAGGGCAACATAGAGCAA 2479  
Qy 1274 AGATCGAGGAGCTCGCAAGCAGCTGTGCGCTGGGCTTCACACCCCGCAAGAAC 1333  
Db 2480 AAATAGAGAGTTAAGAGAACATCTATTAAAGTGGGATTTACCAACCCAGACAAGAAC 2539

Qy 1334 ACCAGAAAGAGCCCGCTTCTGTGTGATGGCTAGAGCTGCACCCCGCAAGTGGACCG 1393  
Db 2540 ATCAGAAAGAACCCCATTTCTTTGGATGGGTATGAATCCATCTCTGACAAATGGACAG 2599  
Qy 1394 TGCAGCCCATAGAGTCCCGAGAAAGAGAGCTGGAACGTGAACAGCATCCAGAAAGCTGG 1453  
Db 2600 TACAGCTTATACAGCTGGCAGAAAAAGATAGCTGACTGTTAATGATATACAGAAGTTAG 2659  
Qy 1454 TGGGCAAGCTGAATCTGGGCCAGCAGATCTACCCCGGCATCAAGTGGCGCGCTGTGCA 1513  
Db 2660 TGGGAAATTAATCTGGGCAAGTCAGATTACGAGGATTAAGTAAGGCACTTTGTA 2719  
Qy 1514 AGCTCTCGCGGCGCCAAAGCCCTGACCGACATCTGTGCCCTGACCGAGAGGCGGAGC 1573  
Db 2720 AACTCTTTAGGGAGCCCAAGCACTTAAACAGACATAGTACCTAATCTGAAGACAGAA 2779  
Qy 1574 TGGACTCGCGCAGAACCGCGAGATCTCTCGCGAGCGCGTGCACGGCTGTACTACGACC 1633  
Db 2780 TAGAATTTGGCAGAGAACAAAGGAAATTTTAAAGAACCCAGTACATGGGGTATATTATGACC 2839  
Qy 1634 CCAGCAAGGACCTGTGTGGCCGAGATCCAGAAAGAGGCGCCACGACAGTGGACCTACCA 1693  
Db 2840 CATCAAAAGACTTGATAGCTGAATATACAGAAACAGGCGATGACCAATGGACATATCAA 2899  
Qy 1694 TCTACAGAGGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1753  
Db 2900 TTTACCGAGAACCATTCAAAATCTGAAAACAGGGAAGTATGCAAAAATAGGACAGGCC 2959  
Qy 1754 ACACCAACGACGTGAAGCAGCTGACCGAGCGCGTGCAGAAAGATCGCCATGAGAGACATCG 1813  
Db 2960 ACATTAATGATGATAAAACAGTTTACAGAGGAGTGCATAAATAATAGCCCTGGAGAGCATAG 3019  
Qy 1814 TGATCTGGGCGAAGACCCCAAGTTCCGCTGCGCCATCCAGAGGAGACCTTGGGAGACCT 1873  
Db 3020 TAAATAGGCGAAGATTTCTTAATTTAGACTACCCATCCAAAAAGAAACATTTGGGAAACAT 3079  
Qy 1874 GGTGACGAGCTACTTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCGCC 1933  
Db 3080 GGTGACAGACTATTTGGCAAGCCACCTGGATTTCTGAGTGGGAGTTTGTATACCCCTC 3139  
Qy 1934 CCTCGTGAAGCTGTGTATCCAGCTGGAGAGAGGACCCATCATCGCGCGCGAGACCTTCT 1993  
Db 3140 TCTTAGTAAATTTATGTTACCGAGCTGGAGAAAGAACCCATAGTAGGAGCAGAAAACCTTCT 3199  
Qy 1994 AGTGCAGCGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGCTACGTGACCGAC 2053  
Db 3200 ATGTAGATGGAGCGCCAAATAGGGAACATAATAGGAAAGCAGGGTATATTACTGACA 3259  
Qy 2054 GGGCGCGCAGAGATCGTGAAGCTGACCGAGACCAACCAAGAGAGCCAGTGCAGG 2113  
Db 3260 GAGGAAGGCAAAAAATTTCTTACTTAACTGAAAACAAATCAAGAGACTGAATTAAGA 3319  
Qy 2114 CCATCCAGCTGGCCCTGAGGACAGCGCAGGAGTGAACATCGTGAACGAGCGCAGT 2173  
Db 3320 CAATTTACTAGCTTTCAAGATTCAGGATCAGAAAGTAAACATAGTAACCTGACTCAGAGT 3379  
Qy 2174 AGCCCTGGGATCATCCAGGCCCGCCAGAGCGAGCGAGCGAGTGGTGAACCCAGA 2233  
Db 3380 ATGCGTTAGGAATCATTCAGACATCCAGATAAGATGAATCAGAGTTAGTCAACCAA 3439  
Qy 2234 TCATCGAGCTGATCAAGAGGAGAGGTTGTTACTGAGCTGGGTGCGCCCGCCACAGG 2293  
Db 3440 TAAATAGAACAAATTAATAAGAGGAAAGGGTCTACCTGTCTATGGGTACCGACATTAAG 3499  
Qy 2294 GCATCGCGGCAACGAGCAGATCGCAAGCTGGTGAAGCAAGGCAATCGCAAGGTGTCT 2353  
Db 3500 GAATTTGGAGGTTAATGAACAGGTAGATAAATTAGTAAGCAAGGGAATCAGGAAGGTCTGT 3559  
Qy 2354 TCCGAGCGCATCGA 2369  
Db 3560 TTCTAGATGGAATAGA 3575

RESULT 12

```

US-09-184-418C-9
; Sequence 9, Application US/09184418C
; Patent No. 6492110
; GENERAL INFORMATION:
; APPLICANT: Hahn, Beatrice
; APPLICANT: Gao, Feng
; APPLICANT: Shaw, Feng
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1
; FILE REFERENCE: D6287
; CURRENT APPLICATION NUMBER: US/09/184,418C
; CURRENT FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 112
; SEQ ID NO 9
; LENGTH: 8972
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: isolates=962M751.3; 137.1632:gag; 1419.4435:pol;
; OTHER INFORMATION: 4380.4958:vif; 4858.5188:vpf; 5169.7814:tat;
; OTHER INFORMATION: 5308.7938:rev; 5407.5667:vpv; 5585.8128:env;
; OTHER INFORMATION: 8130.8753:nef
; US-09-184-418C-9

```

Query Match	48.4%;	Score 1196.2;	DB 4;	Length 8972;
Best local Similarity	70.0%;	Pred. No. 2.6e-189;		
Matches 1662;	Conservative	0;	Mismatches 693;	Indels 20; Gaps 3;
QY	14	TGGCCGAGGCCATGAGCCAGCCACCGCCACATCTCTGATCGAGCGGAGCAACTCA	73	
Db	1214	TGGCTGAGCAATGAGCCAGTAAACAATCAACAATATGATCGAGAAAGCAATTTTA	1273	
QY	74	AGGGCCCCAAGCGGATCATCAAGTGTTCAACTCGGCAAGGAGGGCCACATCGCCGCA	133	
Db	1274	AAGGCCCTAAAGAAATTTGTTAAATGTTTCAACTGTGGCAGGAAGGCATATAGCCAGGA	1333	
QY	134	ACTGCCGCGCCCCCGCAAGAAGGCGTCTCGGAATGCGGCACAGGAGGGCCACCATGTA	193	
Db	1334	ATTTCAGGGCTCTCTGGGAJAJAJAGGCTGTGTGAAATGTGAAAGGAAGGACACCCAATGA	1393	
QY	194	AGGACTGACCGAGCGCCAGCCCAACTCTTTCGCGAGGACCTGGCCCTTCCCCCAGGGCA	253	
Db	1394	AAGACTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAAAATTTGGCCTTCCCAAGGGG	1452	
QY	254	AGCCCCGAGTTCCTCCAGCGAGCG-----AACCGGCCCAACAGCC	295	
Db	1453	AGGCCGGGGAATTCCTTTTCAGAACAGACACAGAGCCAAACGCCCCACAGTCCCAACAGCC	1512	
QY	296	CCACCAGCCGAGCTCAGGTGCGCGCGACAAACCCCGCAGCGAGGCGCGGCGCCGAGC	355	
Db	1513	CCACCAGCAGAGCTTCAGGTTTCGAGGAGACAACCCCTGCCCCGAGCAGGACAGAAA	1572	
QY	356	GCAGGSCACCTGAATCTTCCCCAGATCACTCTGTGGCAGGCCCTCTGGTAGCATCA	415	
Db	1573	GACAAGAAACCTTAATCTGCCCTCAAAATCACTCTTTGGCAGCACCCCTTGTCTCAATAA	1632	
QY	416	AGGTGGCGGCGCAGATCAAGAGGCGCCCTGTGGACACCGCGCGCGACACCGTGCCTGG	475	
Db	1633	AAGTAGGGGTTCAGATAAAGAGGCTCTCTTGGATACAGGACAGATGATACAGTATTAG	1692	
QY	476	AGAGATGAGCCTGCCCCGGAAGTGGAAAGCCCAAGATGATCGGCGGCAATCGGCGGCTTCA	535	
Db	1693	AAGAAATATAATTTGCCAGGAAATGGAAAACAAAAATGATAGGAGGAATTTGGAGGTTTTA	1752	
QY	536	TCAAGGTGCCAGTAGCACAGATCTGTATCGAGATCTCGGCGCAAGAGGCCATCGGCA	595	
Db	1753	TCAAGTAGACAGTATGATCAATACTTATAGNAATTTGTGNAAAAAGGCTATAGGTA	1812	
QY	596	CGGTGCTGATCGGGCCCCACCCCCGTGAACATCATCGGCCGCAACATCTGACCCAGCTGG	655	
Db	1813	CAGTATTAGTAGACCTTACACTCTCAACATAATTGGGAGAAATATCTTGACCCAGCTTG	1872	

Qy	656	GCTGCACCTTGAACTCTCCCATCAGCCCCATCAGAGACCGTGCCTGTGAAGCTGAAGCCCG	715
Db	1873	GCTGCACACTAAATTTTCCAAATTTAGTCTTATTTGAAACTGTACCAAGTAAATTAAGCCAG	1932
Qy	716	GCATGACAGCCGCCAAGGTGAACAGTGGCCCTGTACCGAGGAGAAGATCAAGGCCCTGA	775
Db	1933	GAATGGATGCCCCAAGGGTCAACAAATGGCCATTGACAGAAGAAAATAAAGCATTAA	1992
Qy	776	CCGCCATCTCGAGGAGATGGAGAAGGGCAAGATCACCAAGATCGGCCCGGAGAAC	835
Db	1993	CAGCAATTTGTGAAGAAATGGAAAGAGAGNAAAAATTACAAAAATTTGGCCCTTGAGAA	2052
Qy	836	CCTACAAACCCCGCTGTTCGCCATCAAGAAGAAGCAGCACCAAGTGGCGCAAGCTGG	895
Db	2053	CATATACACTCTCCAGTATTTGCCATAAAAGAAGGACAGTACTAAGTGAGAAAAATTAG	2112
Qy	896	TGNACTTCCCGAGCTGAACAAGCGCACCCAGAGACTTCTGGGAGGTGCAGCTGGGCATCC	955
Db	2113	TAGATTTTCAGGGAACCTCAATTAAGAAGAACTCAGAGCTTTTGGGAAGTTCAATTAGGAATC	2172
Qy	956	CCACCCCGCGCGCTGAAGAAGAAGAGCGTGCAGCTGGAGCGTGGGGGAGACGCGCT	1015
Db	2173	CACACCCAGCGGGTTAAAAAGAAAAAGTCAGTGACAGTACTGGATGTGGGGAGTCGCT	2232
Qy	1016	ACTTCAGCTGCCCCCTGGAGCAGAGACTTTCGGCAAGTATACCGGCTTCACCATCCCAGCA	1075
Db	2233	ATTTTTCAGTTCCTTTAGATTGAAGGCTTCAGGAAATATACTGCATTACCATACCTTAGTA	2292
Qy	1076	TCAACACAGAGACCCCGGCGATCCGCTACCAGTACAACTGCTGCCCGCAGGGCTGGAAG	1135
Db	2293	TAAACATGAACACCTGGGATTAGATATCAATATAATGTCTTCCACAGGATGGAAG	2352
Qy	1136	GCAGCCCCAGACTTCCAGAGCAGATGACCAAGATCCTGGAGGCTTCGCGCGCCGCA	1195
Db	2353	GATCACCATCAATATCCAGAGTAGATGATAAAATCTTTAGAGCCCTTTAGGACACAA	2412
Qy	1196	ACCCGAGATCGTGATCTACAGTACATGACACCTGTACGTGGGAGCGACCTGGAGA	1255
Db	2413	ACCAGAAATAGTTATCTATCAATATATGGATGACTTGTATGTAGGATCTGATTTAGAA	2472
Qy	1256	TCGGCAGCACCGCCAGATCGAGAGCTGGCAGACACCTGCTGCGCTGGGGCTTCA	1315
Db	2473	TAGGGCAACACAGACGCAAAATAGAGAGTTAAGAAACCTTATTGAGATGGGATTTA	2532
Qy	1316	CCACCCCGCACAGAAGCACCAAGAGAGCCCCCTTCTGTGTGGATGGGTACGAGCTGC	1375
Db	2533	CTACACAGACAAGAAGCATCAGAAAGAGCCCCCATTTCTTTGGATGGGTATGAACTCC	2592
Qy	1376	ACCCGACAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGACCCGTGA	1435
Db	2593	ATCTGCAAAATGGACAGTACAGCTTAAAGCTGCAGAAAGAGAGAGCTGACTGTCA	2552
Qy	1436	ACGACATCCAGAAGCTGGTGGGCAAGCTGAATCTGGGCCAGCCAGATCTATCCCCGGCATCA	1495
Db	2653	ATGATATACAGAAGTTAGTGGGAAAAATTAACT-GGCAAGTCAGATTTACCGCAGGGATTA	2711
Qy	1496	AGTGGCCAGCTGTGCAGCTCTCGGGGCGCCAAAGCCCTGACCGCATCGTGCCTCC	1555
Db	2712	AAGTAGGCCAACTGTGTAACTCTTTAGGGAGCCAAAGCATACAGACATAGTACCAT	2771
Qy	1556	TGACCGAGGAGCGAGCTGGAGCTGGCCGAGAACCCGAGATCTCTGGCGAGCCCGGTGC	1615
Db	2772	TGACTGAAGAGGCAGAAATTAGAAATTGGCAGAGCAGGGGAAAATTTCAAAAGAACCACTAC	2831
Qy	1616	ACGCGTGTACTACGACCCCGACAGGACTGGTGGCCGAGATCCAGAGACGAGGCCACG	1675
Db	2832	ATGGAGTATATTATGCCCATCAAAAGACTTAACTAGCTGAAATACAGAAACAAGGGCATG	2891
Qy	1676	ACCAGTGGACCTACAGATCTACCGAGGACCTTCAAGAACCTTGAAGACGCGCAAGTAGC	1735
Db	2892	ACCAATGGACATATCAAGTTTACAAAGAACCAATCAAAATCTGAAAACAGGAAAGTATG	2951

QY 1736 CCAAGATGGCAGCCGCCACACACACACGACGTGAAGCAGCTGACCGAGGCGGTGAGAGA 1795  
DB 2952 CAAAAATGAGGACTGCCACACTATGATGTAAACAGTTAACAGAGCGGTGCANAAAA 3011  
QY 1796 TCGCCATGGAGAGCATCGTGTCTGGGCAAGACCCCAAGTTCGGCTGCCATCCAGA 1855  
DB 3012 TAGCCATGGAAGCATAGTAATATGSGAAGATTCTTAATTTAGGTACCCATTCAA 3071  
QY 1856 AGGAGACCTGGAGACCTGGTGGACCGACTACTGCGAGCCACCTGTGATCCCGAGTGG 1915  
DB 3072 AAGAAACATGGAGACATGGTGGACAGACTATTGGCAAGCCACCTGGATTCTGAGTGG 3131  
QY 1916 AGTTGCGTGAACACCCGCCCTGGTGAAGCTGTGTACACGTGGAGAGAGGCCATCA 1975  
DB 3132 AGTTTGTAACTATCCGCCCTAGTAAATTTATGTACACGTGGAGAAAGAACCCATAG 3191  
QY 1976 TCGGCGCCGAGACCTTTACGTGGACGCGCGGCCAACCGGAGACCAAGATCGGAGG 2035  
DB 3192 CAGGAGCAGAACTTTACTATGTAGATGAGCAGCCAAATAGGGAACCTAAATATGGAAAG 3251  
QY 2036 CCGGCTACGTGACCGACCGCGCGCGCAGAGATCGTGAGCTGACCGAGACCAACCAACC 2095  
DB 3252 CAGGGAATGTTACTGACAGAGAGGCAAAATTTGTACTTAACCTGAACACAAATC 3311  
QY 2096 AGAAGACGAGCTGACAGGCAATCGAGCTGCGCTGACAGACAGCGGACGAGGTGAACA 2155  
DB 3312 AAAAGACTGAATTAACAAGCAATTCAGTTAGCTTTTCAGGATTCAGGATCAAGATTAACA 3371  
QY 2156 TCGTGACCGACAGCAGTACGCTGCGCATCATCCAGCGCCAGCCGACAGAGCGAGA 2215  
DB 3372 TAGTAACAGACTACAGTATGCAATTAGGAATCATCCAGCAACACAGATTAAGATGAAT 3431  
QY 2216 GCGAGCTGGTGAACACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCT 2275  
DB 3432 CAGAATTAGTCAATCAATATAGAACAGTTGATATAAAAGAAAGGGTTTACCTGTAT 3491  
QY 2275 GGGTGCCCGCCACAGGCGCATCGCGCCACAGCAGATCGACAGCTGTGTGACGAGG 2335  
DB 3492 GGGTACCAAGCACAAAGAAATTGGAGGAATGAACAAAGTAGATAAATGTTGAATAGTG 3551  
QY 2336 GCATCCGCAAGGTGTGTCTCTGAGCGCATCGAT 2370  
DB 3552 GAATCAGGAAGTGTGTTTCTAGATGAATAGAT 3586

RESULT 13  
US-09-872-733A-3  
; Sequence 3, Application US/09872733A  
; Patent No. 6656706  
; GENERAL INFORMATION:  
; APPLICANT: The Government of the United States of America, as  
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND  
; FILE REFERENCE: SIV ENV GENES  
; CURRENT APPLICATION NUMBER: US/09/872,733A  
; PRIOR FILING DATE: 2001-06-01  
; PRIOR APPLICATION NUMBER: PCT/US00/34985  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/173,036  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 2467  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Mutated Human  
; OTHER INFORMATION: Immunodeficiency Virus - 1 Pol gene  
US-09-872-733A-3

Query Match 48.2%; Score 1190.6; DB 4; Length 2467;  
Best Local Similarity 84.6%; Pred. No. 1.9e-188;

Matches 1337; Conservative 0; Mismatches 244; Indels 0; Gaps 0;  
QY 790 GAGATGGAAGAGGCGCAAGATCAACCAAGATCGGCCCGAGAAACCCCTTACAAACCCCC 849  
DB 7 GAGATGGAAGAGGCGCAAGATCAACCAAGATCGGCCCGAGAAACCCCTTACAAACCCCA 66  
QY 850 GTGTTCCGCTCAAGAGAGGACAGCACCAGTGGCCAGCTGGTGGACTTCCGCGAG 909  
DB 67 GTCTTCGCAATCAAGAGAGGACAGTACCAAGTGGAGAAAGCTGGTGGACTTTCAGAG 126  
QY 910 CTGAAACAGCGACCCAGGACTTCTGGGAGGTGACCTGGGCATCCCCACCCCGCCG 969  
DB 127 CTGAAACAGGAACTCAGGACTTCTGGGAAATTACAGTGGGCATCCCATCCCGCTGG 186  
QY 970 CTGAAAGAGAGAGCGTGCCTGTGGACGTGGGCGACCTACTTTCAGCTGCTCC 1029  
DB 187 TTGAGAGAGAGAGTCACTGACGTGCTGGATGTGGTGATGCTTCTTCCGTTCC 246  
QY 1030 CTGGACGAGGACTTCGCAAGTACACCGCTTCCATCCCGAGCATCAACAAACGAGACC 1089  
DB 247 TTGGACGAGGACTTCAGGAAAGTACACTGCTTCCATACCTAGCATCAACAGAGACA 306  
QY 1090 CCGGCGATCCGTACCAAGTACAACTGTGCCCCAGGCTGGAAGGCGAGCCCCAGCATC 1149  
DB 307 CCAGGCGATCCGTACCAAGTACAACTGTGCCCCAGGCTGGAAGGCGATCAACAGCCATC 366  
QY 1150 TTCCAGAGCAGCATGACCAAGATCTCGAGCGCTTCCGCGCCCGCAACCCCGAGATCGTG 1209  
DB 367 TTTCAAGCAGCATGACCAAGATCTCGAGCGCTTCCGCAAGCAAAACCCAGCATCGTG 426  
QY 1210 ATCTACAGTACATGAACGACTGTACGTGGGACGCGACCTGGAGATCGGCAGCACCGC 1269  
DB 427 ATCTATCAGTACATGACGACCTCTACGTAGGAACTGACCTGGAGATCGGCAGCACAG 486  
QY 1270 GCCAAGATCGAGGCTGCGCAAGCAGCTGTGCGTGGGGCTTCAACCAACCCCGACAG 1329  
DB 487 ACCAAGATCGAGGCTGAGACAGCATCTGTTGAGTGGGACTGACCAACCAACAGCAAG 546  
QY 1330 AAGCAACAGAGAGCGCCCTTCTGTGGATGGGCTACAGCTGCAACCCCGCAAGTGG 1389  
DB 547 AAGCAACAGAGAACTCCCTTCTGTGGATGGGCTACAGCTGCACTCTGACAGTGG 606  
QY 1390 ACCGTGACGCCATCGAGCTGCCGAGAGGAGCTGGACCGTGAAGACATCCAGAG 1449  
DB 607 ACAGTGCAGCCCATCGTGTGCTGAGAGGACAGCTGAGTGTGAAACGACATACAGAG 666  
QY 1450 CTGCTGGGCAAGCTGAACCTGGGCGCAGAGATCTTACCCTGGCATCAAGGTGGCGAGCTG 1509  
DB 667 CTGCTGGGCAAGTGAACCTGGGCGCAGAGATCTTACCCTGGCATCAAGTGGAGCTG 726  
QY 1510 TGCAGCTGCTGGGCGCGCCAGGCGCTGACCGACATCGTCCCTGACCGAGAGGCC 1569  
DB 727 TGCAGCTGCTTCGAGGAAACCAAGGCACTGACAGAAAGTGTATCCCTGACAGAGAA 786  
QY 1570 GAGCTGAGCTGGCGAGAACCGCGAGATCTTGGCGGAGCGCTGACCGGCTGTACTAC 1629  
DB 787 GAGCTAGACTGGCAGAGAACCGCGAGATCTTGGAGAGCCAGTACATGGAGTGTACTAC 846  
QY 1630 GACCCAGCAAGACCTGTGTGCGCGAGATCCAGAGAGGCGGCACGACGATGGACCTAC 1689  
DB 847 GACCCAGCAAGACCTGTGTGCGCGAGATCCAGAGAGGCGGCAGGCGCAATGGACCTAC 906  
QY 1690 CAGATCTACAGAGCGCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGCAC 1749  
DB 907 CAATCTACAGAGCGCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGAGGGT 966  
QY 1750 GCCACACCAAGAGCTGAAGCAGCTGACCGAGGCGGTGCAAGATGCCATGCCATGAGAGC 1809  
DB 967 GCCACACCAAGAGTGAAGCAGCTGACAGAGGCGAGTGAAGAGATCAACACAGAGAGC 1026  
QY 1810 ATCGTATCTGGGCGAAGACCCCGAGTTCCCGCTCCCATCCAGAGGAGACCTGGAG 1869  
DB 1027 ATCGTATCTGGGCGAAGACTCCCAAGTTCAAAGTGTCCCATCAGAGAGACATGGGAG 1086

1870 ACCTGGTGGACCGACTACTGCGCAGGCGACCTGATCCCGAGTGGGAGTTCGTGAACACC 1929  
1087 ACATGGTGGACCGAGTACTGCGCAGGCGACCTGATCCCTGAGTGGGAGTTCGTGAACACC 1146  
1930 CCCCCCTGTGTAAGCTGTGGTACCAAGCTGGAGGAGGCCATCATCCGCGCGCGAGACC 1989  
1147 CTTCCCTTGTGTAAGCTGTGGTATCAGCTGGAGAGGAAACCCATCTGTTGGGAGCAGAGACC 1206  
1990 TTCTACGTGGACGCGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACC 2049  
1207 TTCTACGTGGATGGGCGAGCCACAGGAGACCAAGCTGGGCAAGCGAGCTACGTGACC 1266  
2050 GACCGGGCGCGCAGAAAGATCGTAGCTGACCGAGACCAACCAAGAGACCGAGCTG 2109  
1267 AACCGAGGACGACAGAAAGTGGTGAACCTGACTGACACCAACCAAGAGACTGAGCTG 1326  
2110 CAGGCCATCAGCTGGCGCTGCGAGGACAGCGGAGAGTGAACATCGTACCGACAGC 2169  
1327 CAGGCCATCTACCTAGCTCTGCAAGACAGCGGAGTGAACATCGTACAGACTCA 1386  
2170 CAGTACGCCCTGGGCATCATCCAGGCGCGACCGCGACCAAGAGCGAGAGCTGTGAAC 2229  
1387 CAGTACGCCATCTGGGCATCATCCAGGACCAACAGACCAATCCGAGTCAGAGCTGTGAAC 1446  
2230 CAGATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACTGAGCTGGGTCCCGCCAC 2289  
1447 CAGATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACTGAGCTGGGTACAGACAC 1506  
2290 AAGGGCATCGCGCGCAACGAGCAGATCGCAAGCTGGTGAAGAGCGCATCCGCAAGGTG 2349  
1507 AAGGAATTGGAGGAATGAACAGTAGATAAATTAGTCAAGTCTGGGATCCGGAAGGTG 1566  
2350 CTGTTCTCTGACCGCATCGAT 2370  
1567 CTGTTCTCTGACCGGATCGAT 1587

RESULT 14  
US-09-184-418C-11  
; Sequence 11, Application US/09184418C  
; Patent No. 6492110  
; GENERAL INFORMATION:  
; APPLICANT: Hahn, Beatrice  
; APPLICANT: Gao, Feng  
; APPLICANT: Shaw, George  
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN  
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1  
; FILE REFERENCE: D6287  
; CURRENT APPLICATION NUMBER: US/09/184,418C  
; CURRENT FILING DATE: 1999-11-02  
; NUMBER OF SEQ ID NOS: 112  
; SEQ ID NO 11  
; LENGTH: 8959  
; TYPE: DNA  
; ORGANISM: Human immunodeficiency virus type 1  
; FEATURE:  
; OTHER INFORMATION: isolates=94IN476.104; 138.1613;"gag";  
; OTHER INFORMATION: 1418.4428;"pol"; 4361.4939;"vif"; 4879.5169;"vpr";  
; OTHER INFORMATION: 5150.7782;"tat"; 5289.7939;"rev"; 5378.5638;"vpu";  
; OTHER INFORMATION: 5556.8129;"env"; 8131.8754;"nef"  
US-09-184-418C-11

Query Match 48.2%; Score 1189; DB 4; Length 8959;  
Best Local Similarity 70.0%; Pred. No. 4e-188;  
Matches 1627; Conservative 0; Mismatches 696; Indels 2; Gaps 2;  
46 AACATCTGTATGACGCGCAGCAATCTCAAGCGCGCGCCCAAGCGGCATCATCAAGTGTCTCAAC 105  
1245 AACATATGATGACGAGAGGCAATTTTAAAGCCCTTAAGAGAAATGTTAAATCTTCAAC 1304  
106 TCGCGAAGAGGCGCCATCGCCCGCAATCGCCGCGCGCCCGCCCGCAAGAGGCGTGTGG 165

1305 TGTGGCAAGGAAGGCGCATAAGCCAGAAATTGAGGGCCCTTAGAAAAAGAGCGCTTTGG 1364  
166 AAGTGGCGGCAAGGAGGCGCCACAGATGAAGGACTGTCACCGAGCGCCAGGCAACTTCTTC 225  
1365 AATGTGGCGCAAGGAGGACACCAATGAAGAGACTGTACTGAGAGCGAGGCTAA-TTTTTT 1423  
226 CGCAGAGACCTGGCTTCCCGCGAGGCGAGCGCGAGTTCGCCAGCGAGGAGAACCGC 285  
1424 AGGAAAAATTGGCGCTTCCCAAGGGGAGCGCAGGAAATTTCTTCAAAACAGGCGCAGA 1483  
286 GCCAACAGCCCCACAGCGCGAGCTGCAAGTGGCGGCGACAAACCCCGCAGCGAGGCC 345  
1484 GCCAACAGCCCCACAGCGAGAGCTTCAGTTCAAGGAGACAAACCCCGCTCCGAAGCA 1543  
346 GCGCGCAGCGCGCAGGCGACCTGAACTTCCCGCAGATCACTGTGGAGCGCCCGCTG 405  
1544 GAGTCTGAAGACAGGGAACCTTAACCTTCCCTCAAAATCACTTTTGGCAGCGACCCCTT 1603  
406 GTGAGCATCAAGGTGGCGCGCAGATCAAGGAGCGCTCTGTGACACCGCGCGCGAGCAG 465  
1604 GTCTCAATAAAGTAGGGGCGCAGATAAAGGAAGCTCTCTTAGACACAGAGCAGATGAT 1663  
466 ACCGTGTGGAGGAGATGAGCTGCCCGCAAGTGGAAAGCCCAAGATGATCGGGCGCATC 525  
1664 ACAGTATTAGAAGAAATAGCTTTTCCAGGAGAGATGGAACCAAAATATGATAGAGGAA 1723  
526 GCGGCTTTCATCAAGGTGGCGCGCAGTACGACCGAGATCTGTGAGATCTCGCGCAAGAAG 585  
1724 GAGGTTTTATCAAGTAGAGCAGATGATCAAACTTATAGAAATTTGTGAAAAAAG 1783  
586 GCCATCGGCGACCGTGTGATCGGCGCCCGCAGCTGAACTCATCGGCGCCCAACATCTG 645  
1784 GCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAGATATGTTG 1843  
646 ACCAGCTGGGCTGCACCTGAACTTCCCATCAGCGCCCATCGAGACCGTGCCTGGAAG 705  
1844 ACTCAGCTTGGATGCACTCTAAATTTTCCAAATTTAGCCCATTTGAACTGTACAGTAA 1903  
706 CTGAAGCGCGCATGCGACGCGCCCAAGGTGAAGCAGTGGCGCCCTGACCGGAGGAAGATC 765  
1904 TTAAGCCAGGAATGGATGGCGCCCAAGGTTAAACAGTGGCCATTGACAGAAGAGAAATA 1963  
766 AAGGCGCTGACCGCATCTGGGAGGAGATGGAGAGGAGGCGCAAGATCACCAAGATCGGC 825  
1964 AAGCATTTAAAGAAATTTGTAAGAAATGGAGAGGAGGAGGAAATTTACAAAATTTGG 2023  
826 CCCGAGAACCCCTCAACACACCGCTGTGCGCATCAAGAGAGGAGGAGCAGCAACAGTGG 885  
2024 CTTGAATTCATATACACTCCAGTATTTGCCATATAAAGGAGGAGCAGTACTAAGTGG 2083  
886 CGCAAGCTGGTGGACTTCGCGGAGCTGAACAGCGGACCCAGGACTTCTGGGAGTGCAG 945  
2084 AGAAAAATTAGTAGATTTTCAAGGAGCTCAATAAAGAACTCAAGACTTTTGGGAGTTCAA 2143  
946 CTGGCATCCCCCAGCGCGCTCAAGAGAGGAGCGGTGACCGTGTGGAGCTG 1005  
2144 TTAGGATACACACCGCAGAGGTTTAAAGAGAAAAATCAGTGACAGTACTGGATGTG 2203  
1006 GCGGACGCTTACTTACGCGTCCCTCGGAGGAGACTTCCGCAAGTACACCGCTTCAAC 1065  
2204 GGGGATGCAATATTTTTCAGTTCTCTTTAGATGAAGGCTTCGGGAAATATCTGCACTTCA 2263  
1066 ATCCCCAGCATCAACAGGAGACCCCGGATCCGCTACCAAGTACCAACGCTGTGCCCGCAG 1125  
2264 ATACCTAGTATTAACATAGAAACACCGAGGATTAAGATATCAATATATGTGTTCCACAG 2323  
1126 GGCTGAAGGAGCGACCCCGAGCATCTTCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTC 1185  
2324 GGATGGAAGGATCACCAGCAATATTCAGAGTAGCATGACAAAATATCTTAGAGCCCTTT 2383  
1186 GCGCGCGCGAACCCCGAGATCGTGTCTACCAAGTACATGGAAGACCTGTAGTGGGAGC 1245  
2384 AGGGCACCAGAAATCCAAAATATAGTCTATCAATATATGATGACTTGTATGTAGGTCT 2443

QY 1246 GACCTGGAGATCGCCGACGACCGCGCCAGATCAGAGAGTGGCAAGCACTGCTCGGC 1305  
DB 2444 GACTTAGAAATAGGCGCATCATAGAGCAAAATAGAGGATTAAAGACATCTATTAAAG 2503  
QY 1306 TGGGCTTTACCAACCCCGCAGCAAGAACCCAGAGAGGAGCCCTTCTTGTGGATGGC 1365  
DB 2504 TGGGATTTACCAACACACAGATAAGAAACATCAGAAAGAACCCCATTTCTTTGGATGGG 2563  
QY 1366 TACGAGCTGCACCCGACCAAGTGCAGCTGCAGCCCATCGAGCTGCCGAGAGGAGC 1425  
DB 2564 TATGAACCTCCATCTCTGACAAATGGACATGAGCTTATAAGCTTAAAGCTGCCAAGAGATGC 2623  
QY 1426 TGGACCGTGAACGACATCAGAAAGCTGTGGGCAAGCTGAACCTGGCCAGCCAGATCTAC 1485  
DB 2624 TGGACTGTCAATGATATACAGAGTTAGTGGGAAATTAACCTGGCAAGTCAGATTAC 2683  
QY 1486 CCCGSCATCAGTCCGCGAGCTGTGCAAGCTGTGCGCGGGCCAGAGCCCTCACCAGC 1545  
DB 2684 CCAGGGATTAAAGTGAAGCACTTTGTAACTCTCTAGGGGGGCCAAAGCACTAACAGAC 2743  
QY 1546 ATCGTGCCCTTGACCGAGGAGCCGAGCTGGAGCTGGCGGAGAACCGGAGATCCTCGGC 1605  
DB 2744 ATAGTACCACCTAACTGAAGAGCAGAAATTAGATTAGCAGAGAACAGGAAATTTCTAAA 2803  
QY 1606 GAGCCCGTGCACGGCGTGTATACGACCCCGCAGCAAGACCTGTGGCGCGAGATCCAGAG 1665  
DB 2804 GAGCCAGTACATGGAGTATATTATGACCCATCAAAAGACTTAATAGCTGAAATACAGAAA 2863  
QY 1666 CAGGCGCCAGCAGCTGACCTTACAGATCTACAGAGGAGCCCTCAAGAACTCAAGACC 1725  
DB 2864 CAGGGCGATGACCAATGACATATCAATTTACCAAGAACCACTTCAAAATCTGAAACA 2923  
QY 1726 GCAAGTACGCAAGATGGCACCGCCCAACAGCAGTGAAGAGCTGACCGAGGCC 1785  
DB 2924 GGAAGTATGCAAAATAGGACTCTCACACTAATGATGTAAGACAGTAAACAGAGGCA 2983  
QY 1786 GTGCAGAGATCGCATGAGAGCATCGTATCGGGCCAGAGACCCCAAGTTGCGCTG 1845  
DB 2984 GTGCAAAATAGCCATAGAAAGCATAGTAATAT -GGGAAGACCCCTTAATTTAGACTA 3042  
QY 1846 CCCATCCAGAGAGACCTGGGAGACCTGTGTGACCGACTAGTGCAGGCCACCTGGATC 1905  
DB 3043 CCCATCCAAAGAAACGTGGGAGACATGTTGACAGACTATGGCAGGCCACCTGGATT 3102  
QY 1906 CCGGATGGAGTTGTTGACACACCCCGCTGTGTGAGCTGTGTGACCTGACCTGGAGAG 1965  
DB 3103 CTTGATGGGAGTTGTTTAATACCCCTCCCTTAGTAAATTAATGATGACCTAGAAATA 3162  
QY 1966 GAGCCCATCATCGGCGCCGAGACCTTTACGTGGACCGCGCGCCCAACCGGAGACCAAG 2025  
DB 3163 GAACCCATAGTAGGAGCAGAACTTTCTATGTAGATGGAGCAGCTAATAGGAAACTAAA 3222  
QY 2026 ATCGGCAAGCGCGGTACTGTGACCGACCGGGCCCGGCGAGAGATGCTGAGCCTGACCGAG 2085  
DB 3223 GTAGGAAAGCAGGGTATGTTACTGACAGAGGAGGCGAATAATTTGTTCTTTAACTGAA 3282  
QY 2086 ACCACCAACAGAGACCGAGCTGCAGGCGCATCTGAGCTGGCCCTGCAGGACAGCGCGAGC 2145  
DB 3283 ACACAAATCAGAGACTGAAATGGACGCAATTCAGCTAGCTTTGCAAGATTCAGGAACA 3342  
QY 2146 GAGGTGAACATGTTGACCGACAGCCAGTACCGCCCTGGGCAATCATCAAGCCCGAGCCGAC 2205  
DB 3343 GAAGTAAACATAGTAACAGACTCACAGTATGCAATTAGGAATCATTTCAAGCACAACCCAGAT 3402  
QY 2206 AAGAGCAGAGGAGCTGTGTGACCAAGATCATCGAGCCTGATCAAGAGGAGAGGTG 2265  
DB 3403 AAAGTGAATCAGAGTTAGTCAACCAATTAATAGAACAAATTAATTAACAAAGAGTTC 3462  
QY 2266 TACCTGAGCTGGTGCCTCCCAACAGGGCATCGGCGCAACGAGCAGATCGACAGCTG 2325  
DB 3463 TATCTGCTG3GTACAGCACATAAAGGAAATTGGAGGGAATGAAACAGTAGATAGATTAA 3522

QY 2326 GTGAGCAAGGGCATCCGCAAGTGCTGTCTCTGGACGCATCGAT 2370  
DB 3523 GTAAGTAGTGGAAATAGGAAAGTACTGTTTCTAGATGGATAGAT 3567  
RESULT 15  
US-09-184-418C-4  
; Sequence 4, Application US/09184418C  
; Patent No. 6492110  
; GENERAL INFORMATION:  
; APPLICANT: Hahn, Beatrice  
; APPLICANT: Gao, Peng  
; APPLICANT: Shaw, George  
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN  
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1  
; FILE REFERENCE: D6287  
; CURRENT APPLICATION NUMBER: US/09/184,418C  
; CURRENT FILING DATE: 1999-11-02  
; NUMBER OF SEQ ID NOS: 112  
; SEQ ID NO 4  
; LENGTH: 8992  
; TYPE: DNA  
; ORGANISM: Human immunodeficiency virus type 1  
; FEATURE:  
; OTHER INFORMATION: isolate=92RW009; 139.1624:gag; 1690.4428:pol(N-terminus uncertain;  
; OTHER INFORMATION: 4373.4951:vif; 4891.5181:vpr; 5162.7801:tat; 5301.7958:rev;  
; OTHER INFORMATION: 5403.5648:vpu; 5566.8146:env; 8150.8773:nef  
US-09-184-418C-4  
Query Match 47.2%; Score 1165.6; DB 4; Length 8992;  
Best Local Similarity 68.9%; Pred. No. 2.9e-184;  
Matches 1627; Conservative 0; Mismatches 729; Indels 4; Gaps 2;  
QY 14 TGGCCGAGGCGCATGAGCCAGGCCACAGCGCCCAACATCTGATGAGCGCAGCAACTTCA 73  
DB 1221 TGGCTGAAGCAATGAGCCAAAGTACAAACCAACAAATATGATGAGAGAGCAATTTTA 1280  
QY 74 AGGCGCCCAAGCGCATCATCAAGTGTTCACCTGGCGCAAGGAGGCCACATCGCCCGCA 133  
DB 1281 AGGCGCAGAGAGAAATTAATAGTGTTCACCTGTGGCAAGAGACACCTAGCCAGAA 1340  
QY 134 ACTGCGCGCGCCCCCGCAAGAGAGGCTGTGAGTGGCGCAAGGAGGCCACAGATGA 193  
DB 1341 ATTGCGAGGCGCCCTAGAAAAAAGGGCTGTGAAATGTGAAAGGAGGAGACCAAAATGA 1400  
QY 194 AGGACTGCACCGAGCCGAGGCCCACTCTTCGCGAGGACCTGGCCTTTCCCGCAGGGCA 253  
DB 1401 AAGACTGCACTGAGACAGGCTAA-TTTTTTAGGAAATTTGGCCTTCCAAAGGGG 1459  
QY 254 AGCCCGCGAGTTCCCGAGCGAGCAACCGCGCCAAAGCCCGCCAGCCAGCCGAGCT-- 311  
DB 1460 AGCCCAAGAAATTTTCCCGAGAGCACTGGAGCCCAACAGCCCGCCACAGCAGAACTTT 1519  
QY 312 -GCAGGTGCGCGGAGCAACCCCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 370  
DB 1520 GGAATGGGGGAGAGATAGCTCTCTCTGAAACAGAGAGAGAGAGAGAGAGAGAGAGAG 1579  
QY 371 ACTTCCCCCAGATCACCCCTGTGGACGCGCCCTGTGTGAGCATCAAGGTGGCGGCGCAGA 430  
DB 1580 ATTCCCTCAAACTACTCTTTGGCAACGACCCCTTGTCAAGTAAATAATAGAGGTGAG 1639  
QY 431 TCAAGAGGCGCTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 490  
DB 1640 TAAAGAGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAGAAATAAATTGC 1699  
QY 491 CCGCAAGTGGAAAGCCCAAGATGATCGCGGSCATCGCGGCTTTCATCAAGGTGCGCGAGT 550  
DB 1700 CAGGAAATGGAACCAAAATGATAGGGGGAATTGGAGGTTTATCAAGGTAAACAGT 1759  
QY 551 ACAGACAGATCTGATCGAGATCTGGGCAAGAGGCCATCGGCGCGCGCGCGCGCGCGCG 610  
DB 1760 ATGATCAAACTACTATAGAAATTTGTGAAAAAAGGCTATAGTACAGTATTAGTAGGAC 1819

Qy 611 CCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTGAAC 670  
Db 1820 CTACATCTGTCAACATTAATGGAGAAATATGTTGACCCAGATTGGTTGTACTTTAACT 1879  
Qy 671 TCCCATCAGCCCATCGAGACCGTGCCTGTAGCTGAGCGCGCGCATGGACGGCCCCA 730  
Db 1880 TTCCAAATTAGTCTTATGAGACTGTACAGTAGCAITTAAGCCAGGAATGGATGGCCCCA 1939  
Qy 731 AGGTGAAGCAGTGCCTGTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGGGAGG 790  
Db 1940 AGGTTAAACAATGCCATTGACAGAGAAATAAAGCAATTAAGAGAAATTTGTACAG 1999  
Qy 791 AGATGGAGAGGGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTACAAACCCCGG 850  
Db 2000 AAATGGAAAGAGGGGAAAATTTCAAAAATCGGGCCTGAAAATCCATATAACACTCCAG 2059  
Qy 851 TGTTTCGCCATCAAGAAGAGCAGCACCAAGTGGCGCAAGCTGGTGGACTTCGCGGAGC 910  
Db 2060 TATTTGCCATAAAGAGAGGACAGTACTAAGTGGAGAAATTAGTAGATTTTCAGGGAAC 2119  
Qy 911 TGAACAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCACTCCGCCACCCCGCGGCC 970  
Db 2120 TCAACAAAGAACTCAAGACTTTTGGGAAGTCCAAATTAGGGATACCAACCCAGAGGGT 2179  
Qy 971 TGAAGAAGAGAGAGCGGTGACCGTGTGACGTGGCGGACGCTACTTCAGCGTGCACC 1030  
Db 2180 TAAAGAGAAAATCAGTGACAGTACTGGATGTGGGGATGCATACTTCTCAGTTCCIT 2239  
Qy 1031 TGGACGAGGACTTCGCAAGTACACCGCTTACCATCCCGAGCATCAACAGAGACCC 1090  
Db 2240 TAGATGAGGCTTCAGGAAATATACTGCATTACCATACCTAGTATAAACAATGAAACAC 2299  
Qy 1091 CCGGCATCCGCTACAGTACAACTGTGCTGCCAGGCTGGAAGGCGAGCCCGCAGATCT 1150  
Db 2300 CAGAAATTAGGTATCAATATATGTGCTTCAAGGATGGAAGATACACGCAATAT 2359  
Qy 1151 TCCAGAGCATGACCAAGATCTCGAGCGCTTCCGCGCCCGCAACCCCGAGATCGTGA 1210  
Db 2360 TCCAAAATAGTATGACAAAATCTTAGAGCCCTTAGGGCACAACAAACCAAGAAATAGTGA 2419  
Qy 1211 TCTACCATGATGACGACCTGTACGTGGGAGCAGCTGGAGATCGGCCAGCAGCGG 1270  
Db 2420 TCTATCAATATAGTATGACTTGTACGTAGATCTGACTTAGAAATAGGCAACATAGAG 2479  
Qy 1271 CCAAGATCGAGAGCTGCGAAGCAGCTGCTGCGCTGGGCTTCAACACCCCGCAAGA 1330  
Db 2480 CAAAAATAGAGAGTTAAGAGAACATCTATTAAAGTGGGGATTTACCACACAGACAAGA 2539  
Qy 1331 AGCACAGAGGAGCCCCCTTCTGTGTGGATGGCTACGAGCTGCACCCCGCAAGTGA 1390  
Db 2540 AACATCAGAAAGAACCTCCATTTCTTTGGATGGGGTATGAACCTTCATCTGCAAAATGGA 2599  
Qy 1391 CCGTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCGTGAACGACATCCAGAAGC 1450  
Db 2600 CAGTACAACCTATACAGCTGCCAGAAAGGATAGCTGGACTGTCAATGATATACAGAGT 2659  
Qy 1451 TGGTGGGCAAGCTGAACTGGGCGAGCGAGATCTACCCCGGATCAAGGTGCGCGAGCTGT 1510  
Db 2660 TAGTGGGAAAATTAACCTGGGCAAGTCAAGTTTATCCAGGGGTAAAGTAAGGCAATTGT 2719  
Qy 1511 GCAAGCTGCTGCGGGGCGCAAGCCCTGACCGGACATCGTGCCTGACCGAGAGGCGG 1570  
Db 2720 GTAACTCTTAGGGGAACCAAGCATTAACAGACATAGTACCACCTAAGTGAAGAGCAG 2779  
Qy 1571 AGCTGGAGCTGGCGAGAACCGCGAGATCTTCGCGGAGCCGCTGCACGGCGTGTATACG 1630  
Db 2780 AATTAGAAATGGCAGAAAACAGGGAAAATTTTAAAGAAACCAAGTACATGGAGTATATTATG 2839  
Qy 1631 ACCCCAGCAAGGACTGTGTGGCCGAGATCCAGAGAGGAGGCGACGACCTGGACCTACC 1690  
Db 2840 ACCCATCAAGAGACTTAATAGCTGAATACAGAAACAGGGGCGATGACCAATGACATATC 2899  
Qy 1691 AGATCTACAGGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACGCCAAGATGCGACCG 1750

Db 2900 AAATTTACCAAGAACCCATTCAAAAATCTGAAAACAGGAAAGTAGTATGCAAAAAGAGGACTG 2959  
Qy 1751 CCCACCAACGACGCTGAAGCAGCTGACCGAGGCGGTGCAGAGATCGCCATCGAGAGCA 1810  
Db 2960 CCCACACTATGACGTAAACAGTTACAGAGGCGAGTGCAAAAGATAGCCATGGAAGCA 3019  
Qy 1811 TCGTGAATCGGGCAAGACCCCAAGTTCGCCCTGCCATCCAGAGAGACCTGGGAGA 1870  
Db 3020 TAGTAATATGGGGAAGACTCTTAATTTAGATTACCCATCCAGAAAAGAACATGGGAAA 3079  
Qy 1871 CCTGTGACCGACTACTTGGCAGGCCACCTGGATCCCGAGTGGGAGTTGTTGAACACCC 1930  
Db 3080 CATGTTGACAGACTATTGGGCAAGCCACCTGGATCTCTGAGTGGAGTTTGTTAATACC 3139  
Qy 1931 CCCCCCTGGTGAAGCTGTGGTACAGCTGGAGAGAGCCCATCATCGGGCCGAGACCT 1990  
Db 3140 CTCCTCTAGTAAATTTATGTTACCAAGTGTAGAGAAAGAACCCCATATTAGGAGCAGAGCTT 3199  
Qy 1991 TCTAGCTGACGGCGCGCCCAACCGGAGACCAAGATCGGCAAGCGCGGTACGTCACCG 2050  
Db 3200 TCTATGTAGATGGAGCAGCTAATCGGNAACTAAATAGGAAAGCAGGGTATGTTACTG 3259  
Qy 2051 ACCGGGCGCGCAGAAGATCGTAGCCTGACCGAGACCAACCAAGAGACCGAGTGC 2110  
Db 3260 ACAGAGGAAGGCAGAAAATTTCTTAACTGAAACAAACAAATCAGAAGACTGAATTAC 3319  
Qy 2111 AGGCCATCCAGCTGGCCCTGCAGGACAGCGGACGAGGTGAACATCGTGACCGACAGCC 2170  
Db 3320 AAGCAATTCAGCTAGCTTTTACAGGATTCAGATCAGAAGTAAACATAGTACAGACTCAC 3379  
Qy 2171 AGTAGCCCTCGGCATCATCCAGGCCCGCAGAGCGAGCGAGCTGGTGAACC 2230  
Db 3380 AGTATGCATTAGGAATCATTCAGCAACACCAGATAGCAGCAATCGGAGCGAGTCAATC 3439  
Qy 2231 AGATCATCGACAGCTGATCAAGAGAGAGGTGTACTGTAGCTGGTGGTCCCGCCACA 2290  
Db 3440 AAATAATAGAACAGTTAATAAAGGAAAGAGTCTACCTGTCTATGGGTACCCAGCACA 3499  
Qy 2291 AGGGCATCGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGC 2350  
Db 3500 AAGGAATGGAGGAAATGAACAAGTAGATAATTAGTAGTGGAAATCAGGAGAGTGC 3559  
Qy 2351 TGTTCCTCGACGGCATCGAT 2370  
Db 3560 TGTTCCTAGATGGAATAGAT 3579

Search completed: April 10, 2004, 16:22:37  
Job time : 129.622 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 12:49:38 ; Search time 594.779 Seconds  
(without alignments)  
15572.780 Million cell updates/sec

Title: US-09-610-313-30  
Perfect score: 2469  
Sequence: 1 gtcagccaccatgcccga.....gggctagcaccgggtgaatcc 2469

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2475585 seqs, 1875730760 residues

Total number of hits satisfying chosen parameters: 4951170

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

ALIGNMENTS

RESULT 1  
US-10-190-435-45  
; Sequence 45, Application US/10190435  
; Publication NO. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: P18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 2457  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt\_C  
US-10-190-435-45

Query Match 99.5%; Score 2457; DB 14; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GCACCATGGCGAGCCATGAGCCAGCCACCGCCACATCTCTGATGAGCGCAGC 66  
Db 1 GCCACCATGGCGAGCCATGAGCCAGCCACCGCCACATCTCTGATGAGCGCAGC 60

QY 67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCACTGCGCAAGAGGCCACATC 126  
Db 61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCACTGCGCAAGAGGCCACATC 120

QY 127 GCCCGCAACTGCCGCGCCCCCGCAAGAGGCGCTCTGGAAGTGTGGCAAGAGGGCCAC 186

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2457	99.5	2457	14	US-10-190-435-45
2	2457	99.5	2457	14	US-10-190-305A-39
3	2434.6	98.6	2457	14	US-10-190-435-44
4	2434.6	98.6	2457	14	US-10-190-305A-38
5	2401.8	97.3	2445	14	US-10-190-435-43
6	2401.8	97.3	2445	14	US-10-190-305A-37
7	2394.8	97.0	3930	14	US-10-190-435-9
8	2393.2	96.9	3930	14	US-10-190-435-10
9	2393.2	96.9	3930	14	US-10-190-435-11
10	2393.2	96.9	5184	14	US-10-190-435-58
11	2393.2	96.9	5184	14	US-10-190-305A-82
12	2362.8	95.7	3531	14	US-10-190-435-13
13	2361.2	95.6	3537	14	US-10-190-435-14
14	2361.2	95.6	3537	14	US-10-190-435-15
15	2360.2	95.6	5145	14	US-10-190-435-12



QY 2347 GTGCTGTTCTTGAGCGGATCGATGCGGCGATCTGTGATCTACAGTACATGAGACCTG 2406  
Db 2341 GTGCTGTTCTTGAGCGGATCGATGCGGCGATCTGTGATCTACAGTACATGAGACCTG 2400  
QY 2407 TACGTGGGACGCGCGCCCTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2463  
Db 2401 TACGTGGGACGCGCGCCCTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

RESULT 2  
US-10-190-305A-39  
; Sequence 39, Application US/10190305A  
; Publication No. US20030198621A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGSEDE, Jan  
; APPLICANT: BARNETT, Susan  
; APPLICANT: LIAN, Ying  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: 2302-18702 / 18702.002  
; CURRENT APPLICATION NUMBER: US/10/190.305A  
; CURRENT FILING DATE: 2002-07-05  
; NUMBER OF SEQ ID NOS: 93  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 39  
; LENGTH: 2457  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt\_C  
US-10-190-305A-39

Query Match 99.5%; Score 2457; DB 14; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GCCACCATGGCGGAGGCGCATGAGCAGGCCACACGCGCCCAACATCTCTGATGACGCGCAGC 66  
Db 1 GCCACCATGGCGGAGGCGCATGAGCAGGCCACACGCGCCCAACATCTCTGATGACGCGCAGC 60  
QY 67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTCGGCAAGAGGGGCCACATC 126  
Db 61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTCGGCAAGAGGGGCCACATC 120  
QY 127 GCCCGCAACTGCGCGCCCGCCCGCAAGAGGGCTGTGGAAGTGGCGGCAAGAGGGCCAC 186  
Db 121 GCCCGCAACTGCGCGCCCGCCCGCAAGAGGGCTGTGGAAGTGGCGGCAAGAGGGCCAC 180  
QY 187 CAGATGAAGGACTGACCGAGCGCCAGCGCAACTTCTTCCGCGAGGACTGCGCTTCCCG 246  
Db 181 CAGATGAAGGACTGACCGAGCGCCAGCGCAACTTCTTCCGCGAGGACTGCGCTTCCCG 240  
QY 247 CAGGCAAGGGCCCGGAGTTCGCCAGCGAGCAGAAACCGCGCCAAACAGCCCCACAGCGGC 306  
Db 241 CAGGCAAGGGCCCGGAGTTCGCCAGCGAGCAGAAACCGCGCCAAACAGCCCCACAGCGGC 300  
QY 307 GAGCTCGAGGTGCGGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGC 366  
Db 301 GAGCTCGAGGTGCGGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGC 360  
QY 367 CTGAACCTTCCCGCAGATCACCTGTGCGAGCGCCCTCTGTGAGCATCAAGGTGGCGGC 426  
Db 361 CTGAACCTTCCCGCAGATCACCTGTGCGAGCGCCCTCTGTGAGCATCAAGGTGGCGGC 420  
QY 427 CAGATCAAGGAGGCGCTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 486  
Db 421 CAGATCAAGGAGGCGCTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 480  
QY 487 CTGCGCGGCAAGTGAAGGCCCAAGATGATCGCGCGCATCGCGGGCTTTCATCAAGGTGCGC 546  
Db 481 CTGCGCGGCAAGTGAAGGCCCAAGATGATCGCGCGCATCGCGGGCTTTCATCAAGGTGCGC 540

QY 547 CAGTACGACCATGCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTCTGATC 606  
Db 541 CAGTACGACCATGCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTCTGATC 600  
QY 607 GSCCCCAACCCCGGTGAACATCATATGSCCGCAACATGTGACCCAGCTGGGTGTCACCTG 666  
Db 601 GSCCCCAACCCCGGTGAACATCATATGSCCGCAACATGTGACCCAGCTGGGTGTCACCTG 660  
QY 667 AACTTCCCATCAGCCCATCGAGACCGTGCCTGTGAAGCTGAAAGCCCGGATGAGAGGC 726  
Db 661 AACTTCCCATCAGCCCATCGAGACCGTGCCTGTGAAGCTGAAAGCCCGGATGAGAGGC 720  
QY 727 CCCAAGGTGAAGCAGTGGCCCTTGACCGAGAGAGAGATCAAGGCCCTGACCGCCATCTGC 786  
Db 721 CCCAAGGTGAAGCAGTGGCCCTTGACCGAGAGAGAGATCAAGGCCCTGACCGCCATCTGC 780  
QY 787 GAGGAGATGGAAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTTACAAACAC 846  
Db 781 GAGGAGATGGAAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTTACAAACAC 840  
QY 847 CCGCTGTTCCCATCAAGAAGAGGAGCAACCAAGTGGCGCAAGCTGGTGGATCTTCGCG 906  
Db 841 CCGCTGTTCCCATCAAGAAGAGGAGCAACCAAGTGGCGCAAGCTGGTGGATCTTCGCG 900  
QY 907 GAGCTGAACAGAGCGACCCAGGACTTCTGGGAGGTGACAGCTGGGATCCCGACCCCGCC 966  
Db 901 GAGCTGAACAGAGCGACCCAGGACTTCTGGGAGGTGACAGCTGGGATCCCGACCCCGCC 960  
QY 967 GGCCTGAAGAGCAAGAGAGCGTGAACCTGTGTGACGTGGGCGAGCGCTTCTTACGCTG 1026  
Db 961 GGCCTGAAGAGCAAGAGAGCGTGAACCTGTGTGACGTGGGCGAGCGCTTCTTACGCTG 1020  
QY 1027 CCGCTGGAGGAGCTTCCGCAAGTACACCGCTTACCCTCCCGCAGCATCAACAAACAG 1086  
Db 1021 CCGCTGGAGGAGCTTCCGCAAGTACACCGCTTACCCTCCCGCAGCATCAACAAACAG 1080  
QY 1087 ACCCCCGGATCCGCTTACCAAGTACAACTGTCTGCCCGCAGGGCTGGAAGGGCAGCCCCAGC 1146  
Db 1081 ACCCCCGGATCCGCTTACCAAGTACAACTGTCTGCCCGCAGGGCTGGAAGGGCAGCCCCAGC 1140  
QY 1147 ATCTTCCAGAGCAGATGACCAAGATCTGTGAGCCCTTCCGCGCGCGCAACCCCGAGATC 1206  
Db 1141 ATCTTCCAGAGCAGATGACCAAGATCTGTGAGCCCTTCCGCGCGCGCAACCCCGAGATC 1200  
QY 1207 GTGATCTACAGTACATGACGACCTGTACGTGGGCTACGAGCTGCGACCCCGACAG 1266  
Db 1201 GTGATCTACAGTACATGACGACCTGTACGTGGGCTACGAGCTGCGACCCCGACAG 1260  
QY 1267 CGCGCAAGATCGAGAGCTGCGCAAGCACTGTGCTGGGTGGGGCTTACCAACCCCGGAC 1326  
Db 1261 CGCGCAAGATCGAGAGCTGCGCAAGCACTGTGCTGGGTGGGGCTTACCAACCCCGGAC 1320  
QY 1327 AAGAAGCAAGAGAGGAGCCCTTCTGTGTGGATGGGCTACGAGCTGCGACCCCGACAG 1386  
Db 1321 AAGAAGCAAGAGAGGAGCCCTTCTGTGTGGATGGGCTACGAGCTGCGACCCCGACAG 1380  
QY 1387 TGGACCGTGCAGCCCATCGAGCTGCGGAGAGAGAGTGGACCGTGAACGACATCTCAG 1446  
Db 1381 TGGACCGTGCAGCCCATCGAGCTGCGGAGAGAGAGTGGACCGTGAACGACATCTCAG 1440  
QY 1447 AAGCTGTGGGCAAGTGAATCTGGGCGCAGCAGATCTACCCCGGATCAAGGTGGCCAG 1506  
Db 1441 AAGCTGTGGGCAAGTGAATCTGGGCGCAGCAGATCTACCCCGGATCAAGGTGGCCAG 1500  
QY 1507 CTGTGCAAGTGTGTGGCGCGCGCCCAAGGGCCCTGACCGACATCTGTGCCCTGTGACCGAGGAG 1566  
Db 1501 CTGTGCAAGTGTGTGGCGCGCGCCCAAGGGCCCTGACCGACATCTGTGCCCTGTGACCGAGGAG 1560  
QY 1567 GCGAGCTGAGCTGGCGGAGAGACCCGAGATCTCTGGCGAGCCCGTGCACCGCGGTGTAC 1626  
Db 1561 GCGAGCTGAGCTGGCGGAGAGACCCGAGATCTCTGGCGAGCCCGTGCACCGCGGTGTAC 1620  
QY 1627 TAGCACCCAGCAAGGACCTGTGTGGCGGAGATCCAGAGCAGGGGCCACGACCAAGTGGAC 1686

Db 1621 TACAGCCCCAGAGACCTGGTGGCCGAGATCCAGAAAGCAGGGCCACGACCAAGTGGACC 1680  
Qy 1687 TACAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746  
Db 1681 TACAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740  
Qy 1747 ACCGCCACACCCACGAGCTGAGAGCTGACGAGCCGCTGAGAGATGCGCATGGAG 1806  
Db 1741 ACCGCCACACCCACGAGCTGAGAGCTGACGAGCCGCTGAGAGATGCGCATGGAG 1800  
Qy 1807 AGCATCGTGTCTGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGGAGACCTGG 1866  
Db 1801 AGCATCGTGTCTGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGGAGACCTGG 1860  
Qy 1867 GAGACCTGGTGGACCGACTACTGCGAGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC 1926  
Db 1861 GAGACCTGGTGGACCGACTACTGCGAGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC 1920  
Qy 1927 ACCGCCCCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGCGCCGAG 1986  
Db 1921 ACCGCCCCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGCGCCGAG 1980  
Qy 1987 ACCTTACGTGAGCGCGGCCAACCGGAGACCAAGATCGGCAAGCGGCTACGTG 2046  
Db 1981 ACCTTACGTGAGCGCGGCCAACCGGAGACCAAGATCGGCAAGCGGCTACGTG 2040  
Qy 2047 ACCGACCGGGCCCGGCAAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2106  
Db 2041 ACCGACCGGGCCCGGCAAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100  
Qy 2107 CTGACGGCATCTACGTGGCCCTTCAGAGACAGCGGACAGAGTGAACATCGTGACCGAC 2166  
Db 2101 CTGACGGCCATCTACGTGGCCCTTCAGAGACAGCGGACAGAGTGAACATCGTGACCGAC 2160  
Qy 2167 AGCCAGTACGCCCTGGGCAATCATCAGGCCCGAGCCGACCAAGAGCGAGCGAGTGGTG 2226  
Db 2161 AGCCAGTACGCCCTGGGCAATCATCAGGCCCGAGCCGACCAAGAGCGAGCGAGTGGTG 2220  
Qy 2227 AACCATATCTGAGCAGCTGATCAAGAGAGAGTGTACTCGTGGTGGCCGCC 2286  
Db 2221 AACCATATCTGAGCAGCTGATCAAGAGAGAGTGTACTCGTGGTGGCCGCC 2280  
Qy 2287 CACAAGGCGATCGGCGGCAAGCAGATCGACAGCTGTGAGCAGGCGCATCCGCAAG 2346  
Db 2281 CACAAGGCGATCGGCGGCAAGCAGATCGACAGCTGTGAGCAGGCGCATCCGCAAG 2340  
Qy 2347 GTGCTGTCTTGGACGCGATCGATGGCGGATCGTATACAGTACATGGACGACCTG 2406  
Db 2341 GTGCTGTCTTGGACGCGATCGATGGCGGATCGTATACAGTACATGGACGACCTG 2400  
Qy 2407 TACGTGGGCGAGCGGCCCTTGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2463  
Db 2401 TACGTGGGCGAGCGGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

RESULT 3  
US-10-435-44  
; Sequence 44, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: P18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10190.435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 44  
; LENGTH: 2457  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM\_C  
US-10-190-435-44

Query Match 98.6%; Score 2434.6; DB 14; Length 2457;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 2453; Conservative 0; Mismatches 4; Indels 6; Gaps 1;  
Qy 1 GTCCAGCGCACCATATGGCCGAGCCATGAGCCAGGCCACCAATCTCTGATCAG 60  
Db 1 GTCCAGCGCACCATATGGCCGAGCCATGAGCCAGGCCACCAATCTCTGATCAG 60  
Qy 61 GCAGCAATCTCAAGGGCCCCCAAGCGCATCATCAAGTCTTCAACTCGGCAAGGGC 120  
Db 61 GCAGCAATCTCAAGGGCCCCCAAGCGCATCATCAAGTCTTCAACTCGGCAAGGGC 120  
Qy 121 CACATCGGCCCAACTCGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGGCAAGG 180  
Db 121 CACATCGGCCCAACTCGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGGCAAGG 180  
Qy 181 GGCACACAGATGAAGGACTGCAACCGAGGCCAGGCCAACTTCTCCGCGAGGACTGGCC 240  
Db 181 GGCACACAGATGAAGGACTGCAACCGAGGCCAGGCCAACTTCTCCGCGAGGACTGGCC 240  
Qy 241 TTCCCGCAGGCAAGGCCCGCGAGTTCCCGCAGCAGAACCGGCCAACAGCCCCACC 300  
Db 241 TTCCCGCAGGCAAGGCCCGCGAGTTCCCGCAGCAGAACCGGCCAACAGCCCCACC 300  
Qy 301 AGCCGAGAGTGCAGGTCGCGCGCAACACCCCGCAGGAGCGCGCGCGCGAGCGCCAG 360  
Db 301 AGCCGAGAGTGCAGGTCGCGCGCAACACCCCGCAGGAGCGCGCGCGCGAGCGCCAG 360  
Qy 361 GGCACCTTGAATCTTCCCGCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGTG 420  
Db 361 GGCACCTTGAATCTTCCCGCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGTG 420  
Qy 421 GCGCGCCAGATCAAGGAGGCCCTGTGTGACACCGCGCGCAGCAGACCGTGTGGAGG 480  
Db 421 GCGCGCCAGATCAAGGAGGCCCTGTGTGACACCGCGCGCAGCAGACCGTGTGGAGG 480  
Qy 481 ATGAGCTTCCCGCGCAGTGAAGCCCAAGATGATCGGCGCATCGGCGGCTTCAATCAAG 540  
Db 481 ATGAGCTTCCCGCGCAGTGAAGCCCAAGATGATCGGCGCATCGGCGGCTTCAATCAAG 540  
Qy 541 GTGCGCCAGTACGACCAAGATCTTGTGAGATCTGCGGCAAGAGGCCATCGGCAACCGTG 600  
Db 541 GTGCGCCAGTACGACCAAGATCTTGTGAGATCTGCGGCAAGAGGCCATCGGCAACCGTG 600  
Qy 601 CTGATCGGCGCCACCGCGTGAAATCATTCGCGCGCAGCATGCTGACCCAGCTGGGTGC 660  
Db 601 CTGATCGGCGCCACCGCGTGAAATCATTCGCGCGCAGCATGCTGACCCAGCTGGGTGC 660  
Qy 661 ACCCTGAACTTCCCGCATCAGGCCCATCGAGACCGTGTGCGGAGAGTGAAGCCCGCATG 720  
Db 661 ACCCTGAACTTCCCGCATCAGGCCCATCGAGACCGTGTGCGGAGAGTGAAGCCCGCATG 720  
Qy 721 GACGCGCCCAAGGTTGAAGCACTGCGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCC 780  
Db 721 GACGCGCCCAAGGTTGAAGCACTGCGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
Qy 781 ATCTCGGAGGAGATGGAAGAGGAGGCAAGATCAACCAAGATCGGCGCGGAGAACCCCTAC 840  
Db 781 ATCTCGGAGGAGATGGAAGAGGAGGCAAGATCAACCAAGATCGGCGCGGAGAACCCCTAC 840  
Qy 841 AACACCCCGTGTTCGCCATCAAGAGAGGAGCAGCAACCAAGTGGCGGCAAGCTGGTGAC 900  
Db 841 AACACCCCGTGTTCGCCATCAAGAGAGGAGCAGCAACCAAGTGGCGGCAAGCTGGTGAC 900  
Qy 901 TTCCGCGAGCTGAACAAGCGCACCGAGGACTTCTGGGAGGTTGAGCTGGGCGATCCCCCAC 960

Db 901 TTCCGCGAGCTTAAACAAGCGCACCCAGGACTTCTGGAGGGTGCAGCTGGGCATCCCCAC 960  
 QY 961 CCCGCGGCTTGAAGAAGAGAGCGTGAACCGTCTGGACGTGGGCGAGCGCTACTTC 1020  
 Db 961 CCCGCGGCTTGAAGAAGAGAGCGTGAACCGTCTGGACGTGGGCGAGCGCTACTTC 1020  
 QY 1021 AGCGTCCGCTGGACGAGGACTTCCGGAAGTACACCGCTTCAACATCCAGCATCAAC 1080  
 Db 1021 AGCGTCCGCTGGACGAGGACTTCCGGAAGTACACCGCTTCAACATCCAGCATCAAC 1080  
 QY 1081 AACGAGACCCCGGCTTCCGCTTACAGTACACGCTGCTGCCCGAGGGCTGGAAGGGCAGC 1140  
 Db 1081 AACGAGACCCCGGCTTCCGCTTACAGTACACGCTGCTGCCCGAGGGCTGGAAGGGCAGC 1140  
 QY 1141 CCAGCATCTTCAGACGAGTACCAAGATCTCTGGAGCGCTTCCGCGCCGCAACCCC 1200  
 Db 1141 CCAGCATCTTCAGACGAGTACCAAGATCTCTGGAGCGCTTCCGCGCCGCAACCCC 1200  
 QY 1201 GAGATCGTATCTACAGTACATGACGACCTGTGAGTGGGCGAGGACCTGAGATCGC 1260  
 Db 1201 GAGATCGTATCTACCA-----GGCCCCCTGTACGTGGGCGAGGACCTGAGATCGC 1260  
 QY 1261 CAGCACCGCGCAAGATCGAGAGTGGCGCAAGCACTGTGCTGGCTGGGGCTTACCAACC 1320  
 Db 1255 CAGCACCGCGCAAGATCGAGAGTGGCGCAAGCACTGTGCTGGCTGGGGCTTACCAACC 1314  
 QY 1321 CCCGACAAGAACCCAGAGAGCCCTTCTCTGTGATGGGTACGAGTGCACCCC 1380  
 Db 1315 CCCGACAAGAACCCAGAGAGCCCTTCTCTGTGATGGGTACGAGTGCACCCC 1374  
 QY 1381 GACAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGGAGCTGGACCGTGAACGAC 1440  
 Db 1375 GACAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGGAGCTGGACCGTGAACGAC 1434  
 QY 1441 ATCCAGAGCTGGTGGGCAAGCTGAACCTGGGCGCCAGCTACCCCGGATCAAGGTG 1500  
 Db 1435 ATCCAGAGCTGGTGGGCAAGCTGAACCTGGGCGCCAGCTACCCCGGATCAAGGTG 1494  
 QY 1501 CGCCAGCTGTCAAGCTGTGCGCGCGCCCAAGGCGCTGACCGACATCGTGCCTGACC 1560  
 Db 1495 CGCCAGCTGTCAAGCTGTGCGCGCGCCCAAGGCGCTGACCGACATCGTGCCTGACC 1554  
 QY 1561 GAGGAGCGAGCTGGAGCTGGCGGAGAACCCGAGATCTCTGGAGCGCGTGCACGCG 1620  
 Db 1555 GAGGAGCGAGCTGGAGCTGGCGGAGAACCCGAGATCTCTGGAGCGCGTGCACGCG 1614  
 QY 1621 GTGTACTAGCACCCAGCAAGGACCTGTGGCGGAGATCCAGAGCGGCGCACGACGAG 1680  
 Db 1615 GTGTACTAGCACCCAGCAAGGACCTGTGGCGGAGATCCAGAGCGGCGCACGACGAG 1674  
 QY 1681 TGGACCTACAGATCTACAGAGCGCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAG 1740  
 Db 1675 TGGACCTACAGATCTACAGAGCGCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAG 1734  
 QY 1741 ATGCGCACCGCCACACCAAGCAGTGAAGCAGCTGACCGAGCGGTGCAGAGATCGCC 1800  
 Db 1735 ATGCGCACCGCCACACCAAGCAGTGAAGCAGCTGACCGAGCGGTGCAGAGATCGCC 1794  
 QY 1801 ATGAGAGCATCTGTATCTGGGCAAGACCCCAAGTTCGCTGCCATCCAGAGAGAG 1860  
 Db 1795 ATGAGAGCATCTGTATCTGGGCAAGACCCCAAGTTCGCTGCCATCCAGAGAGAG 1854  
 QY 1861 ACTGGGAGCTGTGGACCGACTACTGGGACCGCCTGATCCCGAGTGGGAGTTC 1920  
 Db 1855 ACTGGGAGCTGTGGGAGCCGACTACTGGGAGCGCCACCTGATCCCGAGTGGGAGTTC 1914  
 QY 1921 GTGAACACCCCGCTTGTGTGAAGTGTGTGTACCAAGTGGAGAGAGGCCATCATCGGC 1980  
 Db 1915 GTGAACACCCCGCTTGTGTGAAGTGTGTGTACCAAGTGGAGAGAGGCCATCATCGGC 1974  
 QY 1981 GCCGAGACCTTCTACGTGACGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGC 2040

Db 1975 GCCGAGACCTTCTACGTGACCGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGC 2034  
 QY 2041 TACGTGAACGACCGGCGCGGAGAGATCTGTAGCTTACCGAGACCAACCAAGAG 2100  
 Db 2035 TACGTGAACGACCGGCGCGGAGAGATCTGTAGCTTACCGAGACCAACCAAGAG 2094  
 QY 2101 ACCGAGCTGACGAGCCATCCAGCTGGCCCTGACGAGCAGCGGCGAGGCTGAACATCGTG 2160  
 Db 2095 ACCGAGCTGACGAGCCATCCAGCTGGCCCTGACGAGCAGCGGCGAGGCTGAACATCGTG 2154  
 QY 2161 ACCGAGCAGCAGTACGCCCTTGGGCTATCATCAAGGCGGCGGAGCAAGAGCGAGAGCGAG 2220  
 Db 2155 ACCGAGCAGCAGTACGCCCTTGGGCTATCATCAAGGCGGCGGAGCAAGAGCGAGAGCGAG 2214  
 QY 2221 CTGGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGCTGTACCTGAGCTGGTG 2280  
 Db 2215 CTGGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGCTGTACCTGAGCTGGTG 2274  
 QY 2281 CCCGCCCAACAAGGCGATCGCGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGCGATC 2340  
 Db 2275 CCCGCCCAACAAGGCGATCGCGCGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGCGATC 2334  
 QY 2341 CGCAAGGTGTCTTCTGACCGGCTGATGGCGGCTGATCTACCAAGTACATGGAC 2400  
 Db 2335 CGCAAGGTGTCTTCTGACCGGCTGATGGCGGCTGATCTACCAAGTACATGGAC 2394  
 QY 2401 GACCTGTACGTGGCGAGCGGCGGCTTAGGATCGATTAAAGCTTCCCGGGGTAGCACC 2460  
 Db 2395 GACCTGTACGTGGCGAGCGGCGGCTTAGGATCGATTAAAGCTTCCCGGGGTAGCACC 2454  
 QY 2461 GGT 2463  
 Db 2455 GGT 2457

RESULT 4

US-10-190-305A-38  
 ; Sequence 38, Application US/10190305A  
 ; Publication No. US20030198621A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ZUR MEDEDE, Jan  
 ; APPLICANT: BARNETT, Susan  
 ; APPLICANT: LIAN, Ying  
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
 ; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
 ; FILE REFERENCE: 2302-18702 / 18702.002  
 ; CURRENT FILING DATE: 2002-07-05  
 ; NUMBER OF SEQ ID NOS: 93  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 38  
 ; LENGTH: 2457  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM\_C  
 US-10-190-305A-38

Query Match 98.6%; Score 2434.6; DB 14; Length 2457;  
 Best Local Similarity 99.6%; Pred. No. 0;  
 Matches 2453; Conservative 0; Mismatches 4; Indels 6; Gaps 1;  
 QY 1 GTCGACGCCACCATGGCGGAGGCGCATGAGCCAGGCCACCAAGCCCAACATCTGATGCGAG 60  
 Db 1 GTCGACGCCACCATGGCGGAGGCGCATGAGCCAGGCCACCAAGCCCAACATCTGATGCGAG 60  
 QY 61 GCGAGCAACTTCAAGGGCCCAAGCGGCGATCATCAAGTGTCTCAACTGCGGCAAGAGGGC 120  
 Db 61 GCGAGCAACTTCAAGGGCCCAAGCGGCGATCATCAAGTGTCTCAACTGCGGCAAGAGGGC 120  
 QY 121 CACATCCCGCGACCTCCGCGCGCCCGCAAGAGGGCTGCTGGAAGTCCGCAAGAGG 180  
 Db 121 CACATCCCGCGACCTCCGCGCGCCCGCAAGAGGGCTGCTGGAAGTCCGCGCAAGAGG 180

181 GGCCACCAGATGAAGGACTGACACGAGCGCCAGGCAAACTTCTTCCGCGAGGACTGGCC 240 QY  
181 GGCCACCAGATGAAGGACTGACACGAGCGCCAGGCAAACTTCTTCCGCGAGGACTGGCC 240 Db  
241 TTCCCCCAAGGCAAGGCGCGAGTTCCTCCAGCGAGAGAACCGGCGCAACAGCGCCAC 300 QY  
241 TTCCCCCAAGGCAAGGCGCGAGTTCCTCCAGCGAGAGAACCGGCGCAACAGCGCCAC 300 Db  
301 AGCCGCGAGCTGACAGTGGCGGCGACAAACCCCGCAGCGAGCGCGCGCGCGCGAG 360 QY  
301 AGCCGCGAGCTGACAGTGGCGGCGACAAACCCCGCAGCGAGCGCGCGCGCGCGAG 360 Db  
361 GGCACCTGAACTTCCCGCAGATCAACCTGTGTGAGCGCCCTTGTGTGAGGATCAAGGTG 420 QY  
361 GGCACCTGAACTTCCCGCAGATCAACCTGTGTGAGCGCCCTTGTGTGAGGATCAAGGTG 420 Db  
421 GCGCGCAGATCAAGAGGCGCTGTGTGACACCGCGCGCGAGCACCGTGTGTGAGGAG 480 QY  
421 GCGCGCAGATCAAGAGGCGCTGTGTGACACCGCGCGCGAGCACCGTGTGTGAGGAG 480 Db  
481 ATGAGCTGCCCGGCAAGTGAAGGCCCAAGATGATGCGCGGCGATCGCGGGTTTCATCAAG 540 QY  
481 ATGAGCTGCCCGGCAAGTGAAGGCCCAAGATGATGCGCGGCGATCGCGGGTTTCATCAAG 540 Db  
541 GTGCGCGAGTACGACAGATCTGATCGAGATCTGCGCGAGAGAGGCCATCGGACCGGTG 600 QY  
541 GTGCGCGAGTACGACAGATCTGATCGAGATCTGCGCGAGAGAGGCCATCGGACCGGTG 600 Db  
601 CTGATCGGCGCCACCCCGTGAACATCATCGGCGCGCAATGCTGACCCAGCTGGGTGC 660 QY  
601 CTGATCGGCGCCACCCCGTGAACATCATCGGCGCGCAATGCTGACCCAGCTGGGTGC 660 Db  
661 ACCCTGAACTTCCCGCAGATCTGATCGAGATCTGCGCGAGAGAGGCCATCGGACCGGTG 720 QY  
661 ACCCTGAACTTCCCGCAGATCTGATCGAGATCTGCGCGAGAGAGGCCATCGGACCGGTG 720 Db  
721 GACGCGCCCAAGGTAAGCAGTGGCGCCCTGACCGGAGAGAGATCAAGGCGCTGACCGCC 780 QY  
721 GACGCGCCCAAGGTAAGCAGTGGCGCCCTGACCGGAGAGAGATCAAGGCGCTGACCGCC 780 Db  
781 ATCTGGAGGAGATGGAGAGAGAGGCGCGAGATCAACAGATCGGCGCGAGAACCCCTAC 840 QY  
781 ATCTGGAGGAGATGGAGAGAGAGGCGCGAGATCAACAGATCGGCGCGAGAACCCCTAC 840 Db  
841 AACACCCCGCTGTTCGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900 QY  
841 AACACCCCGCTGTTCGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900 Db  
901 TTCGCGAGCTGAACAGCGCAACAGAGTCTTGGAGAGTGCAGCTGGGATCCCGCAC 960 QY  
901 TTCGCGAGCTGAACAGCGCAACAGAGTCTTGGAGAGTGCAGCTGGGATCCCGCAC 960 Db  
961 CCGCGCGCTGAAG 1020 QY  
961 CCGCGCGCTGAAG 1020 Db  
1021 AGCGTCCCTGGAAGAGAGATTCGCGCAAGTACAACCGCTTCAACATCCCGAGCATCAAC 1080 QY  
1021 AGCGTCCCTGGAAGAGAGATTCGCGCAAGTACAACCGCTTCAACATCCCGAGCATCAAC 1080 Db  
1081 AACGAGACCCCGGCTACCGTACAGATACAGTGTGCGCGAGGCTGGAGGGCAGC 1140 QY  
1081 AACGAGACCCCGGCTACCGTACAGATACAGTGTGCGCGAGGCTGGAGGGCAGC 1140 Db  
1141 CCCAGCATTTCCAGAGCAGATGACCAAGATCTTGGAGCCCTTCCGCGCGCGCAACCCC 1200 QY  
1141 CCCAGCATTTCCAGAGCAGATGACCAAGATCTTGGAGCCCTTCCGCGCGCGCAACCCC 1200 Db  
1201 GAGATCGTATACAGTACATGAGACCGTGTAGTGGGAGCGGACCTGGAGATCGG 1260 QY  
1201 GAGATCGTATACAGTACATGAGACCGTGTAGTGGGAGCGGACCTGGAGATCGG 1260 Db  
-----GGCGCCCTGTAGTGGGAGCGGACCTGGAGATCGG 1254

1261 CAGACCGCGCCAAAGATCGAGAGCTGCGCAAGACCTGTGCGTGGGGCTTACACACC 1320 QY  
1261 CAGACCGCGCCAAAGATCGAGAGCTGCGCAAGACCTGTGCGTGGGGCTTACACACC 1320 Db  
1321 CCGGCAAGACACAG 1380 QY  
1321 CCGGCAAGACACAG 1380 Db  
1315 CCGGCAAGAGACACAG 1374 QY  
1315 CCGGCAAGAGACACAG 1374 Db  
1381 GACAACTGAGACCTGACGCCATCGAGCTGCCGAGAGAGAGAGAGAGAGAGAGAGAG 1440 QY  
1381 GACAACTGAGACCTGACGCCATCGAGCTGCCGAGAGAGAGAGAGAGAGAGAGAGAG 1440 Db  
1375 GACAACTGAGACCTGACGCCATCGAGCTGCCGAGAGAGAGAGAGAGAGAGAGAGAG 1434 QY  
1375 GACAACTGAGACCTGACGCCATCGAGCTGCCGAGAGAGAGAGAGAGAGAGAGAGAG 1434 Db  
1441 ATCCAGAGCTGTGGGCAAGCTGAACCTGGGCGACCGAGATCTACCCCGGATCAAGGTG 1500 QY  
1441 ATCCAGAGCTGTGGGCAAGCTGAACCTGGGCGACCGAGATCTACCCCGGATCAAGGTG 1500 Db  
1435 ATCCAGAGCTGTGGGCAAGCTGAACCTGGGCGACCGAGATCTACCCCGGATCAAGGTG 1494 QY  
1435 ATCCAGAGCTGTGGGCAAGCTGAACCTGGGCGACCGAGATCTACCCCGGATCAAGGTG 1494 Db  
1501 CCGAGCTGTGAAGCTGTGCGCGCGCAAGGCGCTGACCGACATCGTGGCCCTGACC 1560 QY  
1501 CCGAGCTGTGAAGCTGTGCGCGCGCAAGGCGCTGACCGACATCGTGGCCCTGACC 1560 Db  
1495 CCGAGCTGTGAAGCTGTGCGCGCGCAAGGCGCTGACCGACATCGTGGCCCTGACC 1554 QY  
1495 CCGAGCTGTGAAGCTGTGCGCGCGCAAGGCGCTGACCGACATCGTGGCCCTGACC 1554 Db  
1561 GAGGAGCGAGCTGGAGCTGGCGGAGAGACCGGAGATCTGCGGAGCGCGTGCACCGC 1620 QY  
1561 GAGGAGCGAGCTGGAGCTGGCGGAGAGACCGGAGATCTGCGGAGCGCGTGCACCGC 1620 Db  
1555 GAGGAGCGAGCTGGAGCTGGCGGAGAGACCGGAGATCTGCGGAGCGCGTGCACCGC 1614 QY  
1555 GAGGAGCGAGCTGGAGCTGGCGGAGAGACCGGAGATCTGCGGAGCGCGTGCACCGC 1614 Db  
1621 GTGTACTACGACCCAGCAAGGACCTGTGTGGCGGAGATCCAGAGCAGGCGCACAGCAG 1680 QY  
1621 GTGTACTACGACCCAGCAAGGACCTGTGTGGCGGAGATCCAGAGCAGGCGCACAGCAG 1680 Db  
1615 GTGTACTACGACCCAGCAAGGACCTGTGTGGCGGAGATCCAGAGCAGGCGCACAGCAG 1674 QY  
1615 GTGTACTACGACCCAGCAAGGACCTGTGTGGCGGAGATCCAGAGCAGGCGCACAGCAG 1674 Db  
1681 TGAGCTACAGATCTACAGAGCGCTTCAAGAGCTTGAAGACCGGCAAGTACCGCAAG 1740 QY  
1681 TGAGCTACAGATCTACAGAGCGCTTCAAGAGCTTGAAGACCGGCAAGTACCGCAAG 1740 Db  
1675 TGAGCTACAGATCTACAGAGCGCTTCAAGAGCTTGAAGACCGGCAAGTACCGCAAG 1734 QY  
1675 TGAGCTACAGATCTACAGAGCGCTTCAAGAGCTTGAAGACCGGCAAGTACCGCAAG 1734 Db  
1741 ATGCGCACCGCCACACCAAGCAGTGAAGCAGCTGACCGAGCGCTGACAGAGATCGCC 1800 QY  
1741 ATGCGCACCGCCACACCAAGCAGTGAAGCAGCTGACCGAGCGCTGACAGAGATCGCC 1800 Db  
1735 ATGCGCACCGCCACACCAAGCAGTGAAGCAGCTGACCGAGCGCTGACAGAGATCGCC 1794 QY  
1735 ATGCGCACCGCCACACCAAGCAGTGAAGCAGCTGACCGAGCGCTGACAGAGATCGCC 1794 Db  
1801 ATGAGAGATCGTGTATCTGGGCGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAG 1860 QY  
1801 ATGAGAGATCGTGTATCTGGGCGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAG 1860 Db  
1795 ATGAGAGATCGTGTATCTGGGCGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAG 1854 QY  
1795 ATGAGAGATCGTGTATCTGGGCGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAG 1854 Db  
1861 ACCTGGAGACCTGTGTGAGCGACTACTGCGAGCGCCACCTGATCCCGAGTGGGAGTTC 1920 QY  
1861 ACCTGGAGACCTGTGTGAGCGACTACTGCGAGCGCCACCTGATCCCGAGTGGGAGTTC 1920 Db  
1855 ACCTGGAGACCTGTGTGAGCGACTACTGCGAGCGCCACCTGATCCCGAGTGGGAGTTC 1914 QY  
1855 ACCTGGAGACCTGTGTGAGCGACTACTGCGAGCGCCACCTGATCCCGAGTGGGAGTTC 1914 Db  
1921 GTGAAACACCCCGCTGTGTGAGCTGTGTGAGCTGTGTGAGAGAGAGAGAGAGAGAG 1980 QY  
1921 GTGAAACACCCCGCTGTGTGAGCTGTGTGAGCTGTGTGAGAGAGAGAGAGAGAGAG 1980 Db  
1915 GTGAAACACCCCGCTGTGTGAGCTGTGTGAGCTGTGTGAGAGAGAGAGAGAGAGAG 1974 QY  
1915 GTGAAACACCCCGCTGTGTGAGCTGTGTGAGCTGTGTGAGAGAGAGAGAGAGAGAG 1974 Db  
1981 GCGGAGACCTTCTACGTGAGCGCGCGCCAAACCGGAGACCAAGATCGGCAAGGCGCGC 2040 QY  
1981 GCGGAGACCTTCTACGTGAGCGCGCGCCAAACCGGAGACCAAGATCGGCAAGGCGCGC 2040 Db  
1975 GCGGAGACCTTCTACGTGAGCGCGCGCCAAACCGGAGACCAAGATCGGCAAGGCGCGC 2034 QY  
1975 GCGGAGACCTTCTACGTGAGCGCGCGCCAAACCGGAGACCAAGATCGGCAAGGCGCGC 2034 Db  
2041 TAGCTGACCGACCGGCGCGCGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAG 2100 QY  
2041 TAGCTGACCGACCGGCGCGCGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAG 2100 Db  
2035 TAGCTGACCGACCGGCGCGCGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAG 2094 QY  
2035 TAGCTGACCGACCGGCGCGCGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAG 2094 Db  
2101 ACCGAGCTGAGGCGCATCCAGTGGCGCTGAGGAGCGGCGAGCGAGTGAACATCGTG 2160 QY  
2101 ACCGAGCTGAGGCGCATCCAGTGGCGCTGAGGAGCGGCGAGCGAGTGAACATCGTG 2160 Db  
2095 ACCGAGCTGAGGCGCATCCAGTGGCGCTGAGGAGCGGCGAGCGAGTGAACATCGTG 2154 QY  
2095 ACCGAGCTGAGGCGCATCCAGTGGCGCTGAGGAGCGGCGAGCGAGTGAACATCGTG 2154 Db  
2161 ACCGAGCGAGTACGCTTGGGCGCATCCAGGCGCGCGAGCGCGCGAGCGAGCGAG 2220 QY  
2161 ACCGAGCGAGTACGCTTGGGCGCATCCAGGCGCGCGAGCGCGCGAGCGAGCGAG 2220 Db  
2155 ACCGAGCGAGTACGCTTGGGCGCATCCAGGCGCGCGAGCGCGCGAGCGAGCGAG 2214 QY  
2155 ACCGAGCGAGTACGCTTGGGCGCATCCAGGCGCGCGAGCGCGCGAGCGAGCGAG 2214 Db  
2221 CTGCTGAGCAGATCATCGAGCGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAG 2280 QY  
2221 CTGCTGAGCAGATCATCGAGCGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAG 2280 Db  
2215 CTGCTGAGCAGATCATCGAGCGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAG 2274 QY  
2215 CTGCTGAGCAGATCATCGAGCGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAG 2274 Db  
2281 CCGCGCCCAAGGCGCATCGGCGCGCAACGAGCAGATCGAAGCTGTGTGAGAGGCGATC 2340 QY  
2281 CCGCGCCCAAGGCGCATCGGCGCGCAACGAGCAGATCGAAGCTGTGTGAGAGGCGATC 2340 Db  
2275 CCGCGCCCAAGGCGCATCGGCGCGCAACGAGCAGATCGAAGCTGTGTGAGAGGCGATC 2334 QY  
2275 CCGCGCCCAAGGCGCATCGGCGCGCAACGAGCAGATCGAAGCTGTGTGAGAGGCGATC 2334 Db  
CGAAAGTGTGTCTCTGGACCGGCGCATCGATGGCGCGCATGTGTATCTACAGTACATGGAC 2400 QY





1549 GCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCCTGCGAGCCCGTGCAGCGGTGAC 1608  
1627 TAGGACCCAGCAAGACCTGGTGGCCGAGATCCAGAGAGCGGCCACGACAGTGGACC 1686  
1609 TAGGACCCAGCAAGACCTGGTGGCCGAGATCCAGAGAGCGGCCACGACAGTGGACC 1668  
1687 TACGAGATCTACGAGAGCCCTTCAAGACCTGAAGACCGGCAAGTACGCAAGATGGC 1746  
1669 TACGAGATCTACGAGAGCCCTTCAAGACCTGAAGACCGGCAAGTACGCAAGATGGC 1728  
1747 ACCGCCACCAACGACGTGAAGCAGTGAACGAGCGCGCTGAGAGATCGCCATGGAG 1806  
1729 ACCGCCACCAACGACGTGAAGCAGTGAACGAGCGCGCTGAGAGATCGCCATGGAG 1788  
1807 AGCATCGTATCTGGGGCAAGACCCCAAGTTCCGCTGCGCCATCCAGAGAGACCTGG 1866  
1789 AGCATCGTATCTGGGGCAAGACCCCAAGTTCCGCTGCGCCATCCAGAGAGACCTGG 1848  
1867 GAGACCTGGTGAACCACTACTGCGAGGCCACTCTGGATCCCGAGTGGAGTTCGTGAAC 1926  
1849 GAGACCTGGTGAACCACTACTGCGAGGCCACTCTGGATCCCGAGTGGAGTTCGTGAAC 1908  
1927 ACCCCCTGCTGTGAAGCTGTGTACAGCTGGAGAGAGCCCATCATCGCGCCCGAG 1986  
1909 ACCCCCTGCTGTGAAGCTGTGTACAGCTGGAGAGAGCCCATCATCGCGCCCGAG 1968  
1987 ACCTTCTAGTGAACGCGCGCCGACCAACCGGAGACCAAGATCGCAAGCGCGCTACGTG 2046  
1969 ACCTTCTAGTGAACGCGCGCCGACCAACCGGAGACCAAGATCGCAAGCGCGCTACGTG 2028  
2047 ACCGACCGGGCGCGGAGAGATCTGTGAGCTGTGACCGAGACCAACCAAGAGACCGAG 2106  
2029 ACCGACCGGGCGCGGAGAGATCTGTGAGCTGTGACCGAGACCAACCAAGAGACCGAG 2088  
2107 CTGACGGCCATCCAGTGGCCCTGCGAGGACAGCGGAGAGAGTGAACATCTGACCGAC 2166  
2089 CTGACGGCCATCCAGTGGCCCTGCGAGGACAGCGGAGAGAGTGAACATCTGACCGAC 2148  
2167 AGCCAGTACGCTGGGCTATCTCCAGCGCCAGCCGACCAAGAGCGAGAGCTGGTG 2226  
2149 AGCCAGTACGCTGGGCTATCTCCAGCGCCAGCCGACCAAGAGCGAGAGCTGGTG 2208  
2227 AACGAGATCATCAGCAGCTGATCAGAGAGGAGAGTGTACTGAGTGGTGGCCGCC 2286  
2209 AACGAGATCATCAGCAGCTGATCAGAGAGGAGAGTGTACTGAGTGGTGGCCGCC 2268  
2287 CACAAGGGCATCGCGCGCAACGAGCAGATCGACAAGCTGTGAGCAAGGGCATCCGCAAG 2346  
2269 CACAAGGGCATCGCGCGCAACGAGCAGATCGACAAGCTGTGAGCAAGGGCATCCGCAAG 2328  
2347 GTGCTGTTCTGAGCGCATCGATGGCGCATCTGATCTACAGTACATGAGACGACCTG 2406  
2329 GTGCTGTTCTGAGCGCATCGATGGCGCATCTGATCTACAGTACATGAGACGACCTG 2388  
2407 TAGCTGGGAGCGCGCGCCCTAGGATCGATTAAAGCTTCCGCGGCTAGCACCCGT 2463  
2389 TAGCTGGGAGCGCGCGCCCTAGGATCGATTAAAGCTTCCGCGGCTAGCACCCGT 2445

## RESULT 6

US-10-190-305A-37  
Sequence 37, Application US/10190305A  
Publication No. US20030198621A1  
GENERAL INFORMATION:  
APPLICANT: ZUR MESEDE, Jan  
APPLICANT: BARNETT, Susan  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
FILE REFERENCE: 2302-18702 / 18702.002  
CURRENT APPLICATION NUMBER: US/10/190,305A  
CURRENT FILING DATE: 2002-07-05

NUMBER OF SEQ ID NOS: 93  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 37  
LENGTH: 2445  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
OTHER INFORMATION: P2Pol.Opt.YMMW\_C  
US-10-190-305A-37

Query Match 97.3%; Score 2401.8; DB 14; Length 2445;

Best Local Similarity 99.2%; Pred. No. 0;  
Matches 2438; Conservative 0; Mismatches 7; Indels 12; Gaps 2;

QY 7 GCCACCATGSCCGAGCGCCATGAGCCAGGCGCACAGCGCCCAACATCTCTGATGAGCGCGCAG 66  
Db 1 GCCACCATGSCCGAGCGCCATGAGCCAGGCGCACAGCGCCCAACATCTCTGATGAGCGCGCAG 60  
QY 67 AACTTCAAGGGCCCCAAGGGCCATCATCAAGTGTTCAACTGCGGCAAGGAGGCGCCATC 126  
Db 61 AACTTCAAGGGCCCCAAGGGCCATCATCAAGTGTTCAACTGCGGCAAGGAGGCGCCATC 120  
QY 127 GCCCGCAACTGCGCGCCCGCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGGAGGCGCCAC 186  
Db 121 GCCCGCAACTGCGCGCCCGCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGGAGGCGCCAC 180  
QY 187 CAGATGAAGACTGACCGAGCGCCAGGCGCAACTTCTTCGCGAGGACTGTGGCTTCCCG 246  
Db 181 CAGATGAAGACTGACCGAGCGCCAGGCGCAACTTCTTCGCGAGGACTGTGGCTTCCCG 240  
QY 247 CAGGGCAAGCGCCCGAGTTCCCGAGGAGCAACCGCGCAACAGCGCCCAACAGCGCCCGC 306  
Db 241 CAGGGCAAGCGCCCGAGTTCCCGAGGAGCAACCGCGCAACAGCGCCCAACAGCGCCCGC 300  
QY 307 GAGCTGAGGTGCGCGCGCAACCCCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGC 366  
Db 301 GAGCTGAGGTGCGCGCGCAACCCCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGC 360  
QY 367 CTGAATTTCCCGCAGATCACCTGTGCGAGCGCCCGCTGTGAGCATCAAGTGGCGGCG 426  
Db 361 CTGAATTTCCCGCAGATCACCTGTGCGAGCGCCCGCTGTGAGCATCAAGTGGCGGCG 420  
QY 427 CAGATCAAGGAGGCGCTGTGGAACACCGCGCGCGAGCAACCGCTGTGAGGAGAGATGAGC 486  
Db 421 CAGATCAAGGAGGCGCTGTGGAACACCGCGCGCGAGCAACCGCTGTGAGGAGAGATGAGC 480  
QY 487 CTGCGCGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGC 546  
Db 481 CTGCGCGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGC 540  
QY 547 CAGTACGACAGATCTCTGATCGAGATCTGGGCAAGAGGCCATCGGCAACCGCTGTGATC 606  
Db 541 CAGTACGACAGATCTCTGATCGAGATCTGGGCAAGAGGCCATCGGCAACCGCTGTGATC 600  
QY 607 GGCGCCACCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCAACCTG 666  
Db 601 GGCGCCACCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCAACCTG 660  
QY 667 AACTTCCCGCATCAGCCCGCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCGCATGGAGCGC 726  
Db 661 AACTTCCCGCATCAGCCCGCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCGCATGGAGCGC 720  
QY 727 CCCAAGGTGAAGCAGTGGCCCGCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGC 786  
Db 721 CCCAAGGTGAAGCAGTGGCCCGCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGC 780  
QY 787 GAGGAGATGGAAGAGGAGGCGCAAGATCACAAGTCCGCCCGCGAGACCCCTACACACC 846  
Db 781 GAGGAGATGGAAGAGGAGGCGCAAGATCACAAGTCCGCCCGCGAGACCCCTACACACC 840  
QY 847 CCCGTGTTCCGCATCAAGAGAGGAGCAGCAACAAAGTGGCGGCAAGCTGTGGACTTCCCG 906

QY	1987	ACCTTCTACGTGGACGGGCGCCGACCAACCGCGAGACCAAGATCGGCAAGGCGCGCTACGTG	2045
Db	1969	ACCTTCTACGTGGACGGGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGCGCTACGTG	2028
QY	2047	ACCGACCGGGCGCGCAGAAAGATCGTAGGCTGACCGAGACCAACCAACAGAAAGCCGAG	2106
Db	2029	ACCGACCGGGCGCGCAGAAAGATCGTAGCCTGACCGAGACCAACCAAGAAAGCCGAG	2088
QY	2107	CTGACGGCCATCAGGTGGCCCTGCAGGACAGCGCAGCGAGGTGAACATCTGTGACCGAC	2166
Db	2089	CTGACGGCCATCAGGTGGCCCTGCAGGACAGCGCAGCGAGGTGAACATCTGTGACCGAC	2148
QY	2167	AGCCAGTAGCCCTGGGCATCATCAGGCCACCGCCGACCAAGACGAGAGCGAGCTGGTG	2226
Db	2149	AGCCAGTAGCCCTGGGCATCATCAGGGCCACGCCGACAGAGCGAGCGAGCTGGTG	2208
QY	2227	AACACGATCATGACGAGCTGATCAAGAGGAGAAAGTGTACTCTGAGCTGGTGGCCGCC	2286
Db	2209	AACACGATCATGACGAGCTGATCAAGAGGAGAAAGTGTACTCTGAGCTGGTGGCCGCC	2268
QY	2287	CACAAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGGTAGCAAGGGCATCCGCAG	2346
Db	2269	CACAAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGGTAGCAAGGGCATCCGCAG	2328
QY	2347	GTGCTGTCTGACGGGCATCGATGGCGGCATCGTATCTACCACTACATGACGACCTG	2406
Db	2329	GTGCTGTCTGACGGGCATCGATGGCGGCATCGTATCTACCACTACATGACGACCTG	2388
QY	2407	TAGCTGGGAGCGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT	2463
Db	2389	TAGCTGGGAGCGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT	2445
RESULT 7			
US-10-190-435-9			
; Sequence 9, Application US/10190435			
; Publication NO. US20030143248A1			
; GENERAL INFORMATION:			
; APPLICANT: ZUR MESEDE, Jan			
; APPLICANT: BARNETT, Susan W.			
; APPLICANT: LIAN, Ying			
; APPLICANT: ENGELBRECHT, Susan			
; APPLICANT: VAN KENSBURG, Estrelita J.			
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C			
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF			
; FILE REFERENCE: PP18133.003 / 2302-18133			
; CURRENT APPLICATION NUMBER: US/10/190,435			
; CURRENT FILING DATE: 2002-12-30			
; NUMBER OF SEQ ID NOS: 319			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 9			
; LENGTH: 3930			
; TYPE: DNA			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence: GagComp_Polmut_C			
US-10-190-435-9			
Query Match 97.0%; Score 2394.8; DB 14; Length 3930;			
Best Local Similarity 99.2%; Pred. No. 0;			
Matches 243; Conservative 0; Mismatches 7; Indels 12; Gaps 2;			
QY	14	TGGCCGAGGCCATGAGCAGGCCACAGCGCCAAACATCTCTGATGACGCGAGCAATTCA	73
Db	1487	TCGCGGAGGCCATGAGCAGGCCACAGCGCCAAACATCTCTGATGACGCGAGCAATTCA	1546
QY	74	AGGGCCCAAGCGCATCATCAAGTGTTCACCTGCGGCAAGGAGGGCCACATCGCCCGCA	133
Db	1547	AGGGCCCAAGCGCATCATCAAGTGTTCACCTGCGGCAAGGAGGGCCACATCGCCCGCA	1606
QY	134	ACTGCGCGCCCCCGCAAGAGGGTGTCTGGAAGTGTGGCGCAAGGAGGGCCACAGATGA	193

1607 ACTGCGCGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCACCAAGATGA 1666  
194 AGGACTGACAGAGCGCGCAGGCAACTTCTCCGAGAGAGCTGGCTTCCCGCAGGCA 253  
1667 AGGACTGCAACCGAGCGCCAGGCAACTTCTTCCGCGAGGAGCTGGCTTCCCGCAGGCA 1726  
254 AGGCGCGGAGTTCCTCCAGCGAGAGAACCGCGCCAAAGAGGCCCAACAGCGCGAGCTGC 313  
1727 AGGCGCGGAGTTCCTCCAGCGAGAGAACCGCGCCAAAGAGGCCCAACAGCGCGAGCTGC 1786  
314 AGGTGCGGCGGACAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
1787 AGGTGCGGCGGACAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846  
374 TCCCCAGATACCCCTGTGGAGCGCCCGCTGTGAGCATCAAGGTGGCGCGCGCGCGCGCG 433  
1847 TCCCCAGATACCCCTGTGGAGCGCCCGCTGTGAGCATCAAGGTGGCGCGCGCGCGCGCG 1906  
434 AGAGGCCCTGTGACACCGCGCGCGAGCACCGTGTGAGGAGATGAGCTGCGCGCG 493  
1907 AGAGGCCCTGTGAGACCGCGCGCGAGCACCGTGTGAGGAGATGAGCTGCGCGCG 1966  
494 GCAAGTGAAGGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAGGTGGCGCGCGCG 553  
1967 GCAAGTGAAGGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAGGTGGCGCGCGCG 2026  
554 ACAGATCTGATCGAGATCTGGGCAAGAGGCCATCGGCAACCGTGTGATCGCGCGCGCA 613  
2027 ACCAGATCTGATCGAGATCTGGGCAAGAGGCCATCGGCAACCGTGTGATCGCGCGCGCA 2086  
614 CCGCGGTGAACATCACTGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTCC 673  
2087 CCGCGGTGAACATCACTGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTCC 2146  
674 CCATAGCCCCATCGAGACCGTGCCTGGAAGCTGGAAGCGCGCGCGCGCGCGCGCGCGCG 733  
2147 CCATAGCCCCATCGAGACCGTGCCTGGAAGCTGGAAGCGCGCGCGCGCGCGCGCGCGCG 2206  
734 TGAAGCAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCATCTCCGAGGAGA 793  
2207 TGAAGCAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCATCTCCGAGGAGA 2266  
794 TGAAGAGGAGGCGCAAGATCAACAGATCGGCGCGCGAGAACCCCTCAACACACCCCGTGT 853  
2267 TGAAGAGGAGGCGCAAGATCAACAGATCGGCGCGCGAGAACCCCTCAACACACCCCGTGT 2326  
854 TCGCCATCAGAGAGAGGAGCAGCACCAAGTGGCGAGAGCTGGTGGACTTCCCGAGCTGA 913  
2327 TCGCCATCAGAGAGAGGAGCAGCACCAAGTGGCGAGAGCTGGTGGACTTCCCGAGCTGA 2386  
914 ACAAGGCGACCCAGGACTTCTGGGAGGTGCGAGCTGGGCGATCCCGCAACCCCGCGCGCTGA 973  
2387 ACAAGGCGACCCAGGACTTCTGGGAGGTGCGAGCTGGGCGATCCCGCAACCCCGCGCGCTGA 2446  
974 AGAAGAGAGAGCGTGAACCGTCTGGAAGTGGCGAGAGCTGGTGGACTTCCCGAGCTGA 1033  
2447 AGAAGAGAGAGCGTGAACCGTCTGGAAGTGGCGAGAGCTGGTGGACTTCCCGAGCTGA 2506  
1034 ACAGGAGCTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAAGAGAGCGCGCG 1093  
2507 ACAGGAGCTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAAGAGAGCGCGCG 2566  
1094 GCATCCGCTACCAAGTACACCGTCTGCGCGAGGCTGGAAGGCGAGCGCGCGAGCTTCC 1153  
2567 GCATCCGCTACCAAGTACACCGTCTGCGCGAGGCTGGAAGGCGAGCGCGCGAGCTTCC 2626  
1154 AGAGCAGCATGACCAAGATCTGGAGCGCTTCCGCGCGCGCGCAACCCCGAGATCGTGTCT 1213  
2627 AGAGCAGCATGACCAAGATCTGGAGCGCTTCCGCGCGCGCGCAACCCCGAGATCGTGTCT 2686  
1214 ACCAGTACATGAGCAGCTGTGAGGCGAGGAGCTGAGAGATCGGCGAGCAGCGCGCGCA 1273  
2687 ACCA-----GGCG 2740

1274 AGATCAGAGAGCTGCGCAAGCACCTGCTGCGTGGGGTTTCAACACCCCGGCAAGAGC 1333  
2741 AGATCAGAGAGCTGCGCAAGCACCTGCTGCGTGGGGTTTCAACACCCCGGCAAGAGC 2800  
1334 ACCAGAGAGAGCGCCCTTCTGATGGGCTAGAGCTGCACCCCGGCAAGTGAACCG 1393  
2801 ACCAGAGAGAGCGCCCTTCTGCCCCAT-----CGAGCTGCACCCCGGCAAGTGAACCG 2854  
1394 TGAGCCCATCGAGCTGCGCGAGAGAGAGTGAACCGTGAACCAATPCAGAGAGCTGG 1453  
2855 TGAGCCCATCGAGCTGCGCGAGAGAGAGTGAACCGTGAACCAATPCAGAGAGCTGG 2914  
1454 TGGGCAAGCTGAACCTGGGCGAGCGAGATCTACCCGGCATCAAGGTGGCGCGAGCTGTGA 1513  
2915 TGGGCAAGCTGAACCTGGGCGAGCGAGATCTACCCGGCATCAAGGTGGCGCGAGCTGTGA 2974  
1514 AGTGTGTGGCGCGCGCAAGAGCGCTGACCGCATCTGTGCCCTGTGACCGAGAGCGAGC 1573  
2975 AGTGTGTGGCGCGCGCAAGAGCGCTGACCGCATCTGTGCCCTGTGACCGAGAGCGAGC 3034  
1574 TGGAGCTGGCGGAGAACCGCGAGATCTCTGGGAGCGCGTGCACCGGTGTACTACGACC 1633  
3035 TGGAGCTGGCGGAGAACCGCGAGATCTCTGGGAGCGCGTGCACCGGTGTACTACGACC 3094  
1634 CCAGAGAGAGCTGTGTGGCGGAGATCCAGAGCAGGGCCACGACCAAGTGAACCTACCA 1693  
3095 CCAGAGAGAGCTGTGTGGCGGAGATCCAGAGCAGGGCCACGACCAAGTGAACCTACCA 3154  
1694 TCTACAGGAGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCAACCGCC 1753  
3155 TCTACAGGAGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCAACCGCC 3214  
1754 ACACCAAGAGCTGGAAGAGCTGACCGAGCGCGTGCAGAGAGATCGCCATGAGAGCATCG 1813  
3215 ACACCAAGAGCTGGAAGAGCTGACCGAGCGCGTGCAGAGAGATCGCCATGAGAGCATCG 3274  
1814 TGATCTGGGCAAGACCGCCCAAGTTCGCGTGCCTTCCATCCAGAGAGGAGATCTGGGAG 1873  
3275 TGATCTGGGCAAGACCGCCCAAGTTCGCGTGCCTTCCATCCAGAGAGGAGATCTGGGAG 3334  
1874 GGTGAGCGGACTACTGCGAGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACACCC 1933  
3335 GGTGAGCGGACTACTGCGAGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACACCC 3394  
1934 CCCTGTGTAAGCTGTGTGTACAGCTGGAGAGAGCGCCATCATCGGCGCGGAGACCTTCT 1993  
3395 CCCTGTGTAAGCTGTGTGTACAGCTGGAGAGAGCGCCATCATCGGCGCGGAGACCTTCT 3454  
1994 ACCTGAGAGCGCGCGCCAAACCGGAGACCAAGATCGGAGAGCGCGCTACGTGACCGACC 2053  
3455 ACCTGAGAGCGCGCGCGCCAAACCGGAGACCAAGATCGGAGAGCGCGCTACGTGACCGACC 3514  
2054 GGGCGCGGCAAGATCTGTGAGCTGTGACCGAGACCAACCAAGAGAGCGAGCTGTGAAC 2113  
3515 GGGCGCGGCAAGATCTGTGAGCTGTGACCGAGACCAACCAAGAGAGCGAGCTGTGAAC 3574  
2114 CCATCCAGCTGGCGCTGCGAGGACCGCGAGCGAGTGAACATCTGAGCGAGCGAGCT 2173  
3575 CCATCCAGCTGGCGCTGCGAGGACCGCGAGCGAGTGAACATCTGAGCGAGCGAGCT 3634  
2174 ACGCCCTGGGCGCATCATCCAGCGCGCGCGCAAGAGAGCGAGCGAGCTGTGAAC 2233  
3635 ACGCCCTGGGCGCATCATCCAGCGCGCGCGCAAGAGAGCGAGCGAGCTGTGAAC 3694  
2234 TCATCGAGAGCTGATCAAGAGAGAGAGTGTACTGAGCTGGTGTCCCGCGCGAGG 2293  
3695 TCATCGAGAGCTGATCAAGAGAGAGAGTGTACTGAGCTGGTGTCCCGCGCGAGG 3754  
2294 GCATCGGCGGCAACGAGCAGATCGAAGAGTGTGTGAGCAAGGGGCGATCCGCAAGGTGTGT 2353  
3755 GCATCGGCGGCAACGAGCAGATCGAAGAGTGTGTGAGCAAGGGGCGATCCGCAAGGTGTGT 3814

QY 2354 TCCTGGACGGCATCGATGGGGCATCTGTGATCTACCACTACATGAGCAAGCTGTACGTGG 2413  
 Db 3815 TCCTGGACGGCATCGATGGGGCATCTGTGATCTACCACTACATGAGCAAGCTGTACGTGG 3874  
 QY 2414 GCAGGGGGCCCTAGATCGATTAAAGCTTCCGGGGCTAGCACCGGT 2463  
 Db 3875 GCAGGGGGCCCTAGATCGATTAAAGCTTCCGGGGCTAGCACCGGT 3924  
 RESULT 8  
 US-10-190-435-10  
 ; Sequence 10, Application US/10190435  
 ; Publication No. US20030143248A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ZUR MEGEDE, Jan  
 ; APPLICANT: BARNETT, Susan W.  
 ; APPLICANT: LIAN, Ying  
 ; APPLICANT: ENGELBRECHT, Susan  
 ; APPLICANT: VAN RENSBURG, Esrelita J.  
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
 ; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
 ; FILE REFERENCE: P18133.003 / 2302-18133  
 ; CURRENT APPLICATION NUMBER: US/10/190/435  
 ; CURRENT FILING DATE: 2002-12-30  
 ; NUMBER OF SEQ ID NOS: 319  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 10  
 ; LENGTH: 3930  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutAtt\_ C  
 US-10-190-435-10  
 Query Match 96.9%; Score 2393.2; DB 14; Length 3930;  
 Best Local Similarity 99.2%; Pred. No. 0;  
 Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;  
 QY 14 TGGCGAGCCATGATGACGAGCCACGAGCCACCATCTGATGAGCGAGCAACTTCA 73  
 Db 1487 TGGCGAGCCATGATGAGCCACGAGCCACCATCTGATGAGCGAGCAACTTCA 1546  
 QY 74 AGGGCCCCAAGCGCATCAAGTCTTCACTGCGGCAAGAGGGCCACATGCGCCGCA 133  
 Db 1547 AGGGCCCCAAGCGCATCAAGTCTTCACTGCGGCAAGAGGGCCACATGCGCCGCA 1606  
 QY 134 ACTGCGGCCCCCGCAGAGAGGCTCTGGAAGTGGCGCAAGAGGGCCACAGATGA 193  
 Db 1607 ACTGCGGCCCCCGCAGAGAGGCTCTGGAAGTGGCGCAAGAGGGCCACAGATGA 1666  
 QY 194 AGGACTGACCGAGCGCCAGGCAACTTCTTCCGCGAGGACCTGGCCCTTCCCGCAGGGCA 253  
 Db 1667 AGGACTGACCGAGCGCCAGGCAACTTCTTCCGCGAGGACCTGGCCCTTCCCGCAGGGCA 1726  
 QY 254 AGGCGCGAGTTCCTGAGCGAGCAAGCCGCGCAACGCGCCCAACGCGCGAGGCTGC 313  
 Db 1727 AGGCGCGAGTTCCTGAGCGAGCAAGCCGCGCAACGCGCCCAACGCGCGAGGCTGC 1786  
 QY 314 AGGTGCGCGGCAACAAACCCCGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
 Db 1787 AGGTGCGCGGCAACAAACCCCGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846  
 QY 374 TCCCCCAGATCACTGTGCGAGCGCCCTCTGATGAGCATCAAGTGGCGCGCGCGCGCG 433  
 Db 1847 TCCCCCAGATCACTGTGCGAGCGCCCTCTGATGAGCATCAAGTGGCGCGCGCGCGCG 1906  
 QY 434 AGGAGCCCTGTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493  
 Db 1907 AGGAGCCCTGTGGACTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1966  
 QY 494 GCAAGTGGAGGCAAGATGATGCGCGCGCATGCGCGGCTTTCATCAAGTGGCGCGCGCG 553  
 Db 1967 GCAAGTGGAGGCAAGATGATGCGCGCGCATGCGCGGCTTTCATCAAGTGGCGCGCGCG 2026

QY 554 ACCAGATCCTGATCGAGATCTGCGCAAGAGGCGCATCGGCAACCGTGTGTATCGGCCCCA 613  
 Db 2027 ACCAGATCCTGATCGAGATCTGCGCAAGAGGCGCATCGGCAACCGTGTGTATCGGCCCCA 2086  
 QY 614 CCCCCGTGACATCATCGGCGCGACATGCTGACCCAGCTGGGCTGACCCCTGAACTTCC 673  
 Db 2087 CCCCCGTGACATCATCGGCGCGACATGCTGACCCAGCTGGGCTGACCCCTGAACTTCC 2146  
 QY 674 CCATCAGCCCCATCGAGACCGTGCCTGTAAGCTGAAAGCCCGGCATGACCGGCCAAGG 733  
 Db 2147 CCATCAGCCCCATCGAGACCGTGCCTGTAAGCTGAAAGCCCGGCATGACCGGCCAAGG 2206  
 QY 734 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCATCTGCGAGGAGA 793  
 Db 2207 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCATCTGCGAGGAGA 2266  
 QY 794 TGGAGAAGGAGGCGCAAGATCACCAAGATCGGCGCGGAGAACCCCTTACAACACCCCGGT 853  
 Db 2267 TGGAGAAGGAGGCGCAAGATCACCAAGATCGGCGCGGAGAACCCCTTACAACACCCCGGT 2326  
 QY 854 TGCCCATCAAG 913  
 Db 2327 TGCCCATCAAG 2386  
 QY 914 ACAAGCGCACCCAGAGATCTTCTGGGAGTGCAGCTGGGCAATCCCGACCGCGCGCTGA 973  
 Db 2387 ACAAGCGCACCCAGAGATCTTCTGGGAGTGCAGCTGGGCAATCCCGACCGCGCGCTGA 2446  
 QY 974 AGAAGAAG 1033  
 Db 2447 AGAAGAAG 2506  
 QY 1034 ACAGAGACTTCGCGAGTACAGCGCTTCAACATCCCGAGCATCAACACAGAGAGAGAGAG 1093  
 Db 2507 ACAGAGACTTCGCGAGTACAGCGCTTCAACATCCCGAGCATCAACACAGAGAGAGAGAG 2566  
 QY 1094 GCATCCGCTACCAAGTACAAAGCTGCTGCCCGCAGGGTGGAGGGCGAGCCCGAGATTTCC 1153  
 Db 2567 GCATCCGCTACCAAGTACAAAGCTGCTGCCCGCAGGGTGGAGGGCGAGCCCGAGATTTCC 2626  
 QY 1154 AGAGAGCATGACCAAGATCTGAGGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1213  
 Db 2627 AGAGAGCATGACCAAGATCTGAGGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 2686  
 QY 1214 ACCAGTACATGAGAGACCTGTACGTGGCGAGGACCTGGAGATCGGCGAGCACCGCGCA 1273  
 Db 2687 ACCA-----GGCCCCCTGTAGTGGCGAGCGACCTGGAGATCGGCGAGCACCGCGCA 2740  
 QY 1274 AGATCAGAGAGTGGCGAAGCACTCTGCGCTGGGCTTCAACACCCCGAGAGAGAGAGAG 1333  
 Db 2741 AGATCAGAGAGTGGCGAAGCACTCTGCGCTGGGCTTCAACACCCCGAGAGAGAGAGAG 2800  
 QY 1334 ACCAG 1393  
 Db 2801 ACCAG 2854  
 QY 1394 TGAG 1453  
 Db 2855 TGAG 2914  
 QY 1454 TGGGAG 1513  
 Db 2915 TGGGAG 2974  
 QY 1514 AGTGTCTGCGAG 1573  
 Db 2975 AGTGTCTGCGAG 3034  
 QY 1574 TGGAGTGGCGAG 1633  
 Db 3035 TGGAGTGGCGAG 3094

1634 CCAGCAAGGACCTGTGGCGGATCCAGNAGCGGCCACGACGAGTGGACCTACCA 1693  
1694 TCTACGAGGACCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGCACCGGCC 1753  
1754 ACACCAACGACGTAAGACGCTACCGAGCGCGTGCAGAGATCGCCATCGAGAGCATCG 1813  
1814 TGATCTGGGCAAGAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGAGACTGGGAGACCT 1873  
1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCGCC 1933  
1934 CCCTGTGAAGCTGTGTACAGCTGGAGAGGAGGCCATCATCGGCGCGGAGACCTTCT 1993  
1994 ACCTGGACGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACC 2053  
2054 GGGCGCGGCAAGATCGTGAGCTGACCGAGACCAACACCAAGAGACCGAGCTCGAG 2113  
2114 CCATCCAGTGGCGCTGCAGGACAGCGGCGAGCGAGTGAACATCGTGAACCGACCGAGT 2173  
2174 ACGCCCTGGGATCATCCAGCGCCAGCGCCAGCAAGAGCGAGCGAGCTGGTGAACCGA 2233  
2234 TCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACTGAGCTGGTGGTGGCGGCCCAAGG 2293  
2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAAGAGGAGGATCCCGCAAGTGTGT 2353  
2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAAGTACATGGACGACCTGTACGTGG 2413  
2414 GCAGCGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 2463  
3095 CCAGCAAGGACCTGTGGCGGAGATCCAGNAGCGGCCACGACGAGTGGACCTACCA 3154  
3155 TCTACGAGGACCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGCACCGGCC 3214  
3155 TCTACGAGGACCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGCACCGGCC 3214  
3275 TGATCTGGGCAAGAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGAGACTGGGAGACCT 3334  
3275 TGATCTGGGCAAGAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGAGACTGGGAGACCT 3334  
3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCGCC 3394  
3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCGCC 3394  
3395 CCCTGTGAAGCTGTGTACAGCTGGAGAGGAGGCCATCATCGGCGCGGAGACCTTCT 3454  
3395 CCCTGTGAAGCTGTGTACAGCTGGAGAGGAGGCCATCATCGGCGCGGAGACCTTCT 3454  
3455 ACCTGGACGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACC 3514  
3455 ACCTGGACGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACC 3514  
3515 GGGCGCGGCAAGATCGTGAGCTGACCGAGACCAACACCAAGAGACCGAGCTCGAG 3574  
3515 GGGCGCGGCAAGATCGTGAGCTGACCGAGACCAACACCAAGAGACCGAGCTCGAG 3574  
3575 CCATCCAGTGGCGCTGCAGGACAGCGGCGAGCGAGTGAACATCGTGAACCGACCGAGT 3634  
3575 CCATCCAGTGGCGCTGCAGGACAGCGGCGAGCGAGTGAACATCGTGAACCGACCGAGT 3634  
3635 ACGCCCTGGGATCATCCAGCGCCAGCGCCAGCAAGAGCGAGCGAGCTGGTGAACCGA 3694  
3635 ACGCCCTGGGATCATCCAGCGCCAGCGCCAGCAAGAGCGAGCGAGCTGGTGAACCGA 3694  
3695 TCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACTGAGCTGGTGGTGGCGGCCCAAGG 3754  
3695 TCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACTGAGCTGGTGGTGGCGGCCCAAGG 3754  
3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAAGAGGAGGATCCCGCAAGTGTGT 3814  
3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAAGAGGAGGATCCCGCAAGTGTGT 3814  
3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAAGTACATGGACGACCTGTACGTGG 3874  
3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAAGTACATGGACGACCTGTACGTGG 3874  
2414 GCAGCGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 2463  
3875 GCAGCGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 3924

RESULT 9  
US-10-435-11

; Sequence 11, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBERG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: PPI18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190.435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 3930  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GagCompPolmutina\_C  
US-10-190-435-11

Query Match 96.98; Score 2393.2; DB 14; Length 3930;  
Best Local Similarity 99.28; Pred. No. 0;  
Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;

QY 14 TGGCCGAGGCGCATGAGCAGGCCACGAGCCCAACATCTCTGATGCGAGCGCAACTTCA 73  
Db 1487 TCGCCGAGGCGCATGAGCAGGCCACGAGCCCAACATCTCTGATGCGAGCGCAACTTCA 1546  
QY 74 AGGCGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGGAGGGCCACATCGCCGCA 133  
Db 1547 AGGCGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGGAGGGCCACATCGCCGCA 1606  
QY 134 ACTGCCGCGCGCCCGCAAGAAGGGTGTCTGGAAGTGCAGCAAGGAGGGCCACAGATGA 193  
Db 1607 ACTGCCGCGCGCCCGCAAGAAGGGTGTCTGGAAGTGCAGCAAGGAGGGCCACAGATGA 1666  
QY 194 AGGACTGACCGAGCGCGCAGGCCCAACTTCTTCCGCGAGGACCTGGCTTCCCGCCAGGCA 253  
Db 1667 AGGACTGACCGAGCGCGCAGGCCCAACTTCTTCCGCGAGGACCTGGCTTCCCGCCAGGCA 1726  
QY 254 AGGCGCGCGAGTTCCTCCAGCGAGCAGAAACCGCGCCCAACAGCCCGCAGCGAGCTGC 313  
Db 1727 AGGCGCGCGAGTTCCTCCAGCGAGCAGAAACCGCGCCCAACAGCCCGCAGCGAGCTGC 1786  
QY 314 AGGTGCGCGCGGCAACACCCCGCAGCGAGCGCGCGCGCGAGCCGAGCGGCGACCTGAACT 373  
Db 1787 AGGTGCGCGCGGCAACACCCCGCAGCGAGCGCGCGCGCGAGCCGAGCGGCGACCTGAACT 1846  
QY 374 TCCCGCAGATCACCTGTGTGCGAGCGCGCCCTGTGTGAGCATCAAGTGGCGGCGCAGATCA 433  
Db 1847 TCCCGCAGATCACCTGTGTGCGAGCGCGCCCTGTGTGAGCATCAAGTGGCGGCGCAGATCA 1906  
QY 434 AGGAGCGCGCTGTGCAACCGCGCGCGAGCAGACCGTGTGTGAGGAGATGAGCGTGGCGG 493  
Db 1907 AGGAGCGCGCTGTGCGCCACCGCGCGCGAGCAGACCGTGTGTGAGGAGATGAGCGTGGCGG 1966  
QY 494 GCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAAGAGTGGCGCAGTACG 553  
Db 1967 GCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAAGAGTGGCGCAGTACG 2026  
QY 554 ACCAGATCCTGATCGAGATCTGGCGCAAGAGGCCATCGGCACCGTGTGATCGGCCCA 613  
Db 2027 ACCAGATCCTGATCGAGATCTGGCGCAAGAGGCCATCGGCACCGTGTGATCGGCCCA 2086  
QY 614 CCCCCGTGAACATCATCGCGCGCAACATCTGACCCAGCTGGGTGACACCTGAACTTCC 673  
Db 2087 CCCCCGTGAACATCATCGCGCGCAACATCTGACCCAGCTGGGTGACACCTGAACTTCC 2146  
QY 674 CCATCAGCGCCCATCGAGACCGTGGCGGTGAAGTGAAGCCCGGATGGAACCGGCCCAAGG 733  
Db 2147 CCATCAGCGCCCATCGAGACCGTGGCGGTGAAGTGAAGCCCGGATGGAACCGGCCCAAGG 2206  
QY 734 TGAACGAGTGGCGCCCTGACCGAGGAGATCAAGGCCCTGACCGCATCTGCGAGGAGA 793  
Db 2207 TGAACGAGTGGCGCCCTGACCGAGGAGATCAAGGCCCTGACCGCATCTGCGAGGAGA 2266  
QY 794 TGGAGAAGGAGGCGCAAGATCACCAAGATCGCGCCCGAGAAACCCCTTACAAACCCCGGTGT 853  
Db 2267 TGGAGAAGGAGGCGCAAGATCACCAAGATCGCGCCCGAGAAACCCCTTACAAACCCCGGTGT 2326  
QY 854 TCGCCATCAAGAGAGGAGCAGCCAGAGTGGCGCGCAAGCTGTGTGACTTCCGCGAGCTGA 913  
Db 2327 TCGCCATCAAGAGAGGAGCAGCCAGAGTGGCGCGCAAGCTGTGTGACTTCCGCGAGCTGA 2386

QY 914 ACAAGCGACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCCACCCCGCGGCTGA 973  
 Db 2387 ACAAGCGACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCCACCCCGCGGCTGA 2446  
 QY 974 AGAAGAAAGAGCGTGCCTGCTGGACGTTGGGAGGCGCTACTTACGCTGCGCCCTGG 1033  
 Db 2447 AGAAGAAAGAGCGTGCCTGCTGGACGTTGGGAGGCGCTACTTACGCTGCGCCCTGG 2506  
 QY 1034 ACGAGGACTTCGGCAAGTACACCGCTTCCACATCCCGAGCATCAACAAACAGAGACCCCG 1093  
 Db 2507 ACGAGGACTTCGGCAAGTACACCGCTTCCACATCCCGAGCATCAACAAACAGAGACCCCG 2566  
 QY 1094 GCATCCGCTACAGTACAGCTGTGCCCCAGGCTGGAAGGCGAGCCCGAGATCTTCC 1153  
 Db 2567 GCATCCGCTACAGTACAGCTGTGCCCCAGGCTGGAAGGCGAGCCCGAGATCTTCC 2626  
 QY 1154 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCTGTCT 1213  
 Db 2627 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCTGTCT 2686  
 QY 1214 ACCAGTACATGAGCACTGTACGTGGGAGGAGCCTGGAGATGGCGAGCACCGCGCA 1273  
 Db 2687 ACCA-----GGCCCCCTGTACGTGGGAGGAGCCTGGAGATGGCGAGCACCGCGCA 2740  
 QY 1274 AGATCGAGGAGTGGCGAAGCACTGTGCGTGGGGTTCACCAACCCCGCAAGAAGC 1333  
 Db 2741 AGATCGAGGAGTGGCGAAGCACTGTGCGTGGGGTTCACCAACCCCGCAAGAAGC 2800  
 QY 1334 ACCAGAAGAGCCCGCTTCTGTGGATGGGGTACAGCTGCGACCCCGCAAGTGGACCG 1393  
 Db 2801 ACCAGAAGAGCCCGCTTCTGTGGATGGGGTACAGCTGCGACCCCGCAAGTGGACCG 2854  
 QY 1394 TSCAGCCATCGAGTGGCCGAGAGGAGAGTGGACCGTGAACGACATCCAGAGCTGG 1453  
 Db 2855 TSCAGCCATCGAGTGGCCGAGAGGAGAGTGGACCGTGAACGACATCCAGAGCTGG 2914  
 QY 1454 TGGGCAAGTGAACCTGGGCGAGCAGATCTACCCCGCATCAAGTGGCGCAGCTGTGCA 1513  
 Db 2915 TGGGCAAGTGAACCTGGGCGAGCAGATCTACCCCGCATCAAGTGGCGCAGCTGTGCA 2974  
 QY 1514 AGCTCTGGCGCGCCAGGCGCTGACCGAGATGTCGCTGCTGACCGAGGAGCGGAGC 1573  
 Db 2975 AGCTCTGGCGCGCCAGGCGCTGACCGAGATGTCGCTGCTGACCGAGGAGCGGAGC 3034  
 QY 1574 TGGAGTGGCGCGAGAACCGGAGATCTCTGCGGAGCCCGTGCACGCGCTGTACTACGAC 1633  
 Db 3035 TGGAGTGGCGCGAGAACCGGAGATCTCTGCGGAGCCCGTGCACGCGCTGTACTACGAC 3094  
 QY 1634 CCAGCAAGGACTGTGGCCGAGATCCAGAAAGAGGCGACGACAGTGGACCTACCGA 1693  
 Db 3095 CCAGCAAGGACTGTGGCCGAGATCCAGAAAGAGGCGACGACAGTGGACCTACCGA 3154  
 QY 1694 TCTACCGAGCGCTTCAGAACCTGAGACCGGAGTACGCCAAGATGGCAGCGGCC 1753  
 Db 3155 TCTACCGAGCGCTTCAGAACCTGAGACCGGAGTACGCCAAGATGGCAGCGGCC 3214  
 QY 1754 ACACCAACGACGTGAAGCAGCTGACCGAGCGCTGCAGAAAGATCGCCATGGAGAGCATCG 1813  
 Db 3215 ACACCAACGACGTGAAGCAGCTGACCGAGCGCTGCAGAAAGATCGCCATGGAGAGCATCG 3274  
 QY 1814 TGATCTGGGGAGAGCCCGCAAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1873  
 Db 3275 TGATCTGGGGAGAGCCCGCAAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3334  
 QY 1874 GTGGAGCGGACTTACTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCGTGAACACCCGCC 1933  
 Db 3335 GTGGAGCGGACTTACTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCGTGAACACCCGCC 3394  
 QY 1934 CCTGTGTGAGTGTGGTACCAAGTGGAGAGGAGCCCATCATCTGGCGCGCGAGACCTTCT 1993  
 Db 3395 CCTGTGTGAGTGTGGTACCAAGTGGAGAGGAGCCCATCATCTGGCGCGCGAGACCTTCT 3454  
 QY 1994 ACGTGGACGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACC 2053

RESULT 10

US-10-190-435-58  
 ; Sequence 58, Application US/10190435  
 ; Publication No. US20030143248A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ZUR MEGEDE, Jan  
 ; APPLICANT: BARNETT, Susan W.  
 ; APPLICANT: LIAN, Ying  
 ; APPLICANT: ENGELBRECHT, Susan  
 ; APPLICANT: VAN RENSBURG, Estrelita J.  
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
 ; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF  
 ; FILE REFERENCE: P18133.003 / 2302-18133  
 ; CURRENT APPLICATION NUMBER: US/10/190,435  
 ; CURRENT FILING DATE: 2002-12-30  
 ; NUMBER OF SEQ ID NOS: 319  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 58  
 ; LENGTH: 5184  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: TatRevNefgagCpolIna C  
 US-10-190-435-58

Query Match 96.9%; Score 2393.2; DB 14; Length 5184;  
 Best Local Similarity 99.2%; Pred. No. 0;  
 Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;  
 QY 14 TGGCCGAGGCGCATGAGCCAGGCCACCGAGCCCAACATCTCTGATGCGAGCGCAGCACTTCA 73  
 Db 2741 TCGCCGAGGCGCATGAGCCAGGCCACCGAGGCCCAACATCTCTGATGCGAGCGCAGCACTTCA 2800  
 QY 74 AGGGCCCCAAGCGCATCATCAAGTGTTCAACTGCGCGCAGGAGGCGCCACATCGCCCGCA 133  
 Db 2801 AGGGCCCCAAGCGCATCATCAAGTGTTCAACTGCGCGCAGGAGGCGCCACATCGCCCGCA 2860  
 QY 134 ACTGCGCGCGCCCGCGCAAGAGGGCTGTCTGGAAGTGGCGCAAGAGGGCGCCACAGATGA 193  
 Db 2861 ACTGCGCGCGCCCGCGCAAGAGGGCTGTCTGGAAGTGGCGCAAGAGGGCGCCACAGATGA 2920



Qy 194 AGAATGACCGAGCGCAGGCAACTTCTTCGCGAGGACCTGGGCTTCCCGGAGGCA 253  
Db 2921 AGGACTGACCGAGCGCAGGCAACTTCTTCGCGAGGACCTGGGCTTCCCGGAGGCA 2980  
Qy 254 AGGCGCGGAGTTCCCGAGCGAGCAGAAACCGCGCAACAGCCGCCACAGCGCGAGCTGC 313  
Db 2981 AGGCGCGGAGTTCCCGAGCGAGCAGAAACCGCGCAACAGCCGCCACAGCGCGAGCTGC 3040  
Qy 314 AGGTGCGCGGAGCAACCCCGCGAGCGAGCGCGCGCGAGCGCGCAAGGCAACCTGAACT 373  
Db 3041 AGGTGCGCGGAGCAACCCCGCGAGCGAGCGCGCGCGAGCGCGCAAGGCAACCTGAACT 3100  
Qy 374 TCCCGCAGATCACCTGTGCGAGCGCCCTGTGTGAGCATCAAGGTGGGCGGCGAGATCA 433  
Db 3101 TCCCGCAGATCACCTGTGCGAGCGCCCTGTGTGAGCATCAAGGTGGGCGGCGAGATCA 3160  
Qy 434 AGGAGGCCCTGCTGGAACACCGCGCGCGAGCGCGCGCGAGCGCGCAAGGCAACCTGAACT 493  
Db 3161 AGGAGGCCCTGCTGGAACACCGCGCGCGAGCGCGCGCGAGCGCGCAAGGCAACCTGAACT 3220  
Qy 494 GCAAGTGGAGGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGGCGGAGTACG 553  
Db 3221 GCAAGTGGAGGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGGCGGAGTACG 3280  
Qy 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAGGCCATCGGCAACCGTGTGATCGGCCCCA 613  
Db 3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAGGCCATCGGCAACCGTGTGATCGGCCCCA 3340  
Qy 614 CCCCCTGGAACATCATCGGCGCGAATGTCGACCCAGCTGGGCTGACCCCTGAATTC 673  
Db 3341 CCCCCTGGAACATCATCGGCGCGAATGTCGACCCAGCTGGGCTGACCCCTGAATTC 3400  
Qy 674 CCATCAGCCCCATCGAGACCGTGCCTGMAAGTGAAGCCCGGCGATGAGCGGCCCCAAGG 733  
Db 3401 CCATCAGCCCCATCGAGACCGTGCCTGMAAGTGAAGCCCGGCGATGAGCGGCCCCAAGG 3460  
Qy 734 TGAAGCATGTCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGCGAGAGA 793  
Db 3461 TGAAGCATGTCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGCGAGAGA 3520  
Qy 794 TGGAGAGGAGGCAAGATCAACAGATCGGCCCGGAGAACCCCTACAAACCCCGCTGT 853  
Db 3521 TGGAGAGGAGGCAAGATCAACAGATCGGCCCGGAGAACCCCTACAAACCCCGCTGT 3580  
Qy 854 TCGGCATCAAG 913  
Db 3581 TCGGCATCAAG 3640  
Qy 914 ACAAGCGCACCGAGACTTCTGGAGGTGACGTGGGATCCCGCCACCGCGCGGCTGA 973  
Db 3641 ACAAGCGCACCGAGACTTCTGGAGGTGACGTGGGATCCCGCCACCGCGCGGCTGA 3700  
Qy 974 AGAAGAGAGAGCGTGACCGTGCTGGAAGTGGGCGAGCGGCTACTTCAGCGTGGCGCTGG 1033  
Db 3701 AGAAGAGAGAGCGTGACCGTGCTGGAAGTGGGCGAGCGGCTACTTCAGCGTGGCGCTGG 3760  
Qy 1034 ACGAGGACTTCGCAAGTACACCGCTTACCATCCCGAGCATCAACAAACGAGACCCCG 1093  
Db 3761 ACGAGGACTTCGCAAGTACACCGCTTACCATCCCGAGCATCAACAAACGAGACCCCG 3820  
Qy 1094 GCATCCGCTACAGTACAACTGTGCGCCCGAGGCTGGAAGGCGAGCGCCAGCATCTTC 1153  
Db 3821 GCATCCGCTACAGTACAACTGTGCGCCCGAGGCTGGAAGGCGAGCGCCAGCATCTTC 3880  
Qy 1154 AGAGCAGCATGACCAAGATCTGGAGCCCTTCGCGCGCCGCAACCCCGAGATCGTGATCT 1213  
Db 3881 AGAGCAGCATGACCAAGATCTGGAGCCCTTCGCGCGCCGCAACCCCGAGATCGTGATCT 3940  
Qy 1214 ACCAGTACATGAGCAGCTGTACGTGGGCGAGCACTTGGAGATGGCGGAGCGCGCCCA 1273  
Db 3941 ACCA-----GGCCCCCTGTACGTGGGCGAGCACTTGGAGATGGCGGAGCGCGCCCA 3994  
Qy 1274 AGATCGAGGAGCTGGCGAAGCACTGTGCGCTGGGCTTCACCAACCCCGAGAGAGC 1333

Db 3995 AGATCGAGGAGCTGGCGAAGCACTGTGCTGGCTGGGCTTCAACACCCCGACAAGAGC 4054  
Qy 1334 ACCAAGAGGAGCCCCCTTCTGTGGATGGGCTACGAGTGTGACCCCGACAAGTGAACG 1393  
Db 4055 ACCAAGAGGAGCCCCCTTCTGTGGCTTCTGTGGCTTCTGTGGCTTCTGTGGCTTCTGTGG 4108  
Qy 1394 TGGAGCCATCGAGCTGCCCGAGAGGAGAGTGGACCGTGAACGATCCAGAGTGTG 1453  
Db 4109 TGGAGCCATCGAGCTGCCCGAGAGGAGAGTGGACCGTGAACGATCCAGAGTGTG 4168  
Qy 1454 TGGGCAAGCTGAATGGGCGGCGAGATCTACCCCGGATCAAGGTGGCGGAGTGTGCA 1513  
Db 4169 TGGGCAAGCTGAATGGGCGGCGAGATCTACCCCGGATCAAGGTGGCGGAGTGTGCA 4228  
Qy 1514 AGCTGCTGGGCGGCGCGCAAGGCGCTGACCGACATCTGTGCGCTTACCGAGAGCGGAGC 1573  
Db 4229 AGCTGCTGGGCGGCGCGCAAGGCGCTTACCGGACATCTGTGCGCTTACCGAGAGCGGAGC 4288  
Qy 1574 TGGAGCTGGCGGAGAAACCGCGAGATCTGTGCGGAGCGCGCTGCAAGGCTGTACTACGAC 1633  
Db 4289 TGGAGCTGGCGGAGAAACCGCGAGATCTGTGCGGAGCGCGCTGCAAGGCTGTACTACGAC 4348  
Qy 1634 CCAGCAAGGACTGGTGGCGGAGATCCAGAGCAAGGCGCGAGGAGTGGAGTGGAGTGGAGTGG 1693  
Db 4349 CCAGCAAGGACTGGTGGCGGAGATCCAGAGCAAGGCGCGAGGAGTGGAGTGGAGTGGAGTGG 4408  
Qy 1694 TCTACAGGAGCGCTTCAAGAACTGAAGACCGGCAAGTACGCGCAAGTGGCGACCGCC 1753  
Db 4409 TCTACAGGAGCGCTTCAAGAACTGAAGACCGGCAAGTACGCGCAAGTGGCGACCGCC 4468  
Qy 1754 ACACCAAGCGAGTGAAGCAGCTGACCGAGCGCTGCAAGAGATCCCATGGAGAGATCG 1813  
Db 4469 ACACCAAGCGAGTGAAGCAGCTGACCGAGCGCTGCAAGAGATCCCATGGAGAGATCG 4528  
Qy 1814 TGATCTGGGCGAAGACCCCAAGTTTCGCGCTGCGCCATCCAGAGGAGACCTGGGAGACCT 1873  
Db 4529 TGATCTGGGCGAAGACCCCAAGTTTCGCGCTGCGCCATCCAGAGGAGACCTGGGAGACCT 4588  
Qy 1874 GGTGAGCGACTACTGCGAGCGCACTGTGATTCGCGAGTGGGAGTTCGTGAACACCCCG 1933  
Db 4589 GGTGAGCGACTACTGCGAGCGCACTGTGATTCGCGAGTGGGAGTTCGTGAACACCCCG 4648  
Qy 1934 CCCTGGTGAAGTGTGTGATCCAGCTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCT 1993  
Db 4649 CCCTGGTGAAGTGTGTGATCCAGCTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCT 4708  
Qy 1994 ACGTGGACGGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGGCTAGCTGACCGAGC 2053  
Db 4709 ACGTGGACGGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGGCTAGCTGACCGAGC 4768  
Qy 2054 GGGGCGGCGAGAGATCTGTGAGCTTACCGAGACCAACCAAGAGAGACCGAGTGGAGC 2113  
Db 4769 GGGGCGGCGAGAGATCTGTGAGCTTACCGAGACCAACCAAGAGAGACCGAGTGGAGC 4828  
Qy 2114 CCATCCAGTGGCGCTGAGGAGACGCGGAGCGGAGCGGAGGAGTGAACATCGTGAACCGAGC 2173  
Db 4829 CCATCCAGTGGCGCTGAGGAGACGCGGAGCGGAGCGGAGGAGTGAACATCGTGAACCGAGC 4888  
Qy 2174 ACGCCTGGGATCATCAGGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2233  
Db 4889 ACGCCTGGGATCATCAGGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4948  
Qy 2234 TCATCGAGCAGCTGATCAAGAGAGAGAGAGTGTACCTGAGCTGGTGGCGCGCCACAGG 2293  
Db 4949 TCATCGAGCAGCTGATCAAGAGAGAGAGAGTGTACCTGAGCTGGTGGCGCGCCACAGG 5008  
Qy 2294 GCATCGGCGGCAACGAGCAGATTCGAAGAGTGGTGGAGCAAGGAGTTCGCAAGTGTGT 2353  
Db 5009 GCATCGGCGGCAACGAGCAGATTCGAAGAGTGGTGGAGCAAGGAGTTCGCAAGTGTGT 5068  
Qy 2354 TCTGAGCGGATCGATGGCGGATCTGATCTACAGTACATGAGACCACTGTACGTTGG 2413



Db 5069 TCCTGGACGGCATCGATGGGGCATCGTGTACTTACCAGTACATGAGACCTGTACGTGG 5128  
QY 2414 GCAGGGGGCCCTAGGATCGATTAAAGCTTCCGGGGCTAGCACCGGT 2463  
Db 5129 GCAGGGGGCCCTAGGATCGATTAAAGCTTCCGGGGCTAGCACCGGT 5178

RESULT 11

US-10-190-305A-82  
; Sequence 82; Application US/10190305A  
; Publication No. US20030198621A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan  
; APPLICANT: LIAN, Ying  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
; FILE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: 2302-18702 / 18702.002  
; CURRENT APPLICATION NUMBER: US/10/190.305A  
; NUMBER OF SEQ ID NOS: 93  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 82  
; LENGTH: 5184  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: TatRevNefgagCpolIna C  
US-10-190-305A-82

Query Match 96.9%; Score 2393.2; DB 14; Length 5184;  
Best Local Similarity 99.2%; Pred. No. 0;  
Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;  
QY 14 TGGCCGAGGCCATGAGCCAGCCACCAGCGCCACATCTCTGATGACGCGGAGGCAATTTCA 73  
Db 2741 TCGCGAGGCCATGAGCCAGCCACCAGCGCCACATCTCTGATGACGCGGAGGCAATTTCA 2800  
QY 74 AGGGCCCCAAGCGATCATCAAGTGTCTTCACTCGCGGAGGAGGCGCAATGCCCGCA 133  
Db 2801 AGGGCCCCAAGCGATCATCAAGTGTCTTCACTCGCGGAGGAGGCGCAATGCCCGCA 2860  
QY 134 ACTCGCGCGCCCCCGCAAGAGGCTGCTGGAAGTGGCGGAGGAGGCGCCACAGATGA 193  
Db 2861 ACTCGCGCGCCCCCGCAAGAGGCTGCTGGAAGTGGCGGAGGAGGCGCCACAGATGA 2920  
QY 194 AGGACTGCAACCGAGGCGGAGGCGCAACTTTTCCGCGAGGACCTTGGCCTTCCCGCAGGCA 253  
Db 2921 AGGACTGCAACCGAGGCGGAGGCGCAACTTTTCCGCGAGGACCTTGGCCTTCCCGCAGGCA 2980  
QY 254 AGGCGCGGAGTTCGCCAGCGAGCAGAACCGCGCCACAGCCCAACAGCGCGGAGGTGC 313  
Db 2981 AGGCGCGGAGTTCGCCAGCGAGCAGAACCGCGCCACAGCCCAACAGCGCGGAGGTGC 3040  
QY 314 AGGTGCGCGGCGCAAAACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
Db 3041 AGGTGCGCGGCGCAAAACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3100  
QY 374 TCCCCCAGATCACCTGTGGCAGCGCCCTTGTGAGTCAAGTGGCGGCGCGAGTCA 433  
Db 3101 TCCCCCAGATCACCTGTGGCAGCGCCCTTGTGAGTCAAGTGGCGGCGCGAGTCA 3160  
QY 434 AGGAGGCGCTCTGGACACCGGCGCGCGAGCAGACCGTGTGGAGGAGATGAGCGTGGCCG 493  
Db 3161 AGGAGGCGCTCTGGACACCGGCGCGCGAGCAGACCGTGTGGAGGAGATGAGCGTGGCCG 3220  
QY 494 GCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTCAATCAAGTGGCGGCGAGTACG 553  
Db 3221 GCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTCAATCAAGTGGCGGCGAGTACG 3280  
QY 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCGCGCGCGCGCGCGCGCGCG 613

Db 3281 ACCAGATCCTGATCGAGATCTGGCGCAAGAGGCCATCGGCACCGTGTGTGCGGCCCA 3340  
QY 614 CCCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCAACCTGAACTTCC 673  
Db 3341 CCCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCAACCTGAACTTCC 3400  
QY 674 CCATCAGCCCCCATCGAGACCGGTGCGCGTGAAGCTGAAGCCCGCGCATGAGACGCCCAAG 733  
Db 3401 CCATCAGCCCCCATCGAGACCGGTGCGCGTGAAGCTGAAGCCCGCGCATGAGACGCCCAAG 3460  
QY 734 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGGAGGA 793  
Db 3461 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGGAGGA 3520  
QY 794 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGGAGGA 853  
Db 3521 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGGAGGA 3580  
QY 854 TGCATCAGAAAGAAAGAGAGACCAAGTGGCGCAAGCTGTGTGAGCTTCCGCGAGCTGA 913  
Db 3581 TGCATCAGAAAGAAAGAGAGACCAAGTGGCGCAAGCTGTGTGAGCTTCCGCGAGCTGA 3640  
QY 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGATCCCCCAGCCCGCGGCTGA 973  
Db 3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGATCCCCCAGCCCGGCTGA 3700  
QY 974 AGAAGAAGAGAGCTGACCGTGTGGAGCGTGGCGAGCGCTTACTTCAAGCTGCCCCCTGG 1033  
Db 3701 AGAAGAAGAGAGCTGACCGTGTGGAGCGTGGCGAGCGCTTACTTCAAGCTGCCCCCTGG 3760  
QY 1034 AGGAGGACTTCCGCAAGTACACCGCTTACCATCCCCAGCATCAACAGAGACCCCGC 1093  
Db 3761 AGGAGGACTTCCGCAAGTACACCGCTTACCATCCCCAGCATCAACAGAGACCCCGC 3820  
QY 1094 GCATCCGCTACCAAGTACAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1153  
Db 3821 GCATCCGCTACCAAGTACAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3880  
QY 1154 AGAGCAGATGACCAAGATCTGAGAGCGCTTCCGCGCGCGCAACCCCGAGATCGTGATCT 1213  
Db 3881 AGAGCAGATGACCAAGATCTGAGAGCGCTTCCGCGCGCGCAACCCCGAGATCGTGATCT 3940  
QY 1214 ACCAGTACATGAGACGACCTGTAGTGGGAGCGACTTGGAGATCGGCGAGCACCGCGCA 1273  
Db 3941 ACCA-----GGCCCCCTGTAGTGGGAGCGACTTGGAGATCGGCGAGCACCGCGCA 3994  
QY 1274 AGATCGAGAGCTGGCAAGACCTGTGCTGGCTGGGGCTTCAACACCCCGCAAGAGAC 1333  
Db 3995 AGATCGAGAGCTGGCAAGACCTGTGCTGGCTGGGGCTTCAACACCCCGCAAGAGAC 4054  
QY 1334 ACCAGAGGAGCGCCCTTCTGTGGATGGGCTACGAGCTGCACCCCGCAAGTGGAGCG 1393  
Db 4055 ACCAGAGGAGCGCCCTTCTGTGGATGGGCTACGAGCTGCACCCCGCAAGTGGAGCG 4108  
QY 1394 TGCAGCCCATCGAGCTGCGCGAGAGAGAGTGGACCGTGAACGACATCCAGAGCTGG 1453  
Db 4109 TGCAGCCCATCGAGCTGCGCGAGAGAGAGTGGACCGTGAACGACATCCAGAGCTGG 4168  
QY 1454 TGGGCAAGCTGAACCTGGGCGAGCAGATCTACCCCGCATCAAGTGGCGCGAGCTGTGCA 1513  
Db 4169 TGGGCAAGCTGAACCTGGGCGAGCAGATCTACCCCGCATCAAGTGGCGCGAGCTGTGCA 4228  
QY 1514 AGCTGCTGCGCGCGCCCAAGGCTTACCGCATCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1573  
Db 4229 AGCTGCTGCGCGCGCCCAAGGCTTACCGCATCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4288  
QY 1574 TGGAGCTGGCGAGAACCGCGAGATCTTGGCGGAGCGCGTGCAGCGGCTGTACTACGACC 1633  
Db 4289 TGGAGCTGGCGAGAACCGCGAGATCTTGGCGGAGCGCGTGCAGCGGCTGTACTACGACC 4348  
QY 1634 CCAGCAAGGACCTGTGGCGGAGATCCAGAAAGAGGCGCCAGAGCCAGTGGAGCTTACACAGA 1693  
Db 4349 CCAGCAAGGACCTGTGGCGGAGATCCAGAAAGAGGCGCCAGAGCCAGTGGAGCTTACACAGA 4408

Query Match	95.7%	Score 2362.8	DB 14	Length 3531
Best Local Similarity	98.4%	Pred. No. 0		
Matches 2411	Conservative 0	Mismatches 27	Indels 12	Gaps 2
QY 14	TGGCCGAGGCATGAGCCATAGCCAGGCCACAGGCCCAACATCTCTATGTCAGCGGAGCAACTTCA	73		
DB 1088	TGGCCGAGGCATGAGCCAGGCCACACAGCGTGATGATCAGAGAGCAACTTTAAAA	1147		
QY 74	AGSGCCCAAGCGCATCATCAAGTGTCTCAACTGCGCAAGAGGGGCCACATCGCCGCA	133		
DB 1148	AGSGCCCAAGCGCATCATCAAGTGTCTCACTGCGCAAGAGGGGCCACATCGCCGCA	1207		
QY 134	ACTGCCGCGCCCCGCCAAGAAAGGCTGTGTGAATGTCGGCAAGAGGGGCCACAGATGA	193		
DB 1208	ACTGCCGCGCCCCGCCAAGAAAGGCTGTGTGAATGTCGGCAAGAGGGGCCACAGATGA	1267		
QY 194	AGGACTGCACCGAGCGCCAGGCCAACTTCTCCGCGAGGACTGTGCCCTTCCCCAGGGCA	253		
DB 1268	AGGACTGCACCGAGCGCCAGGCCAACTTCTCCGCGAGGACTGTGCCCTTCCCCAGGGCA	1327		
QY 254	AGGCCCCGCGAGTTCCCGACGAGCAGAAACCGGCCCAACAGCCCCACAGCCGCGAGTGC	313		
DB 1328	AGGCCCCGCGAGTTCCCGACGAGCAGAAACCGGCCCAACAGCCCCACAGCCGCGAGTGC	1387		
QY 314	AGGTGCGCGCGCAACAACCCCGCAGCGAGCCGCGCCAGCGCCAGCGCACCTGTAACT	373		
DB 1388	AGGTGCGCGCGCAACAACCCCGCAGCGAGCCGCGCCAGCGCCAGCGCACCTGTAACT	1447		
QY 374	TCCCCCAGATACCCCTGTGTGACGCGCCCTGTGTAGCATCAAGGTGGCGGCCAGATCA	433		
DB 1448	TCCCCCAGATACCCCTGTGTGACGCGCCCTGTGTAGCATCAAGGTGGCGGCCAGATCA	1507		
QY 434	AGGAGGCCCTGCTGGACAACCGCGCCGACACACCGTGTGTGAGAGATGAGCTGCCCG	493		
DB 1508	AGGAGGCCCTGCTGGACAACCGCGCCGACACACCGTGTGTGAGAGATGAGCTGCCCG	1567		
QY 494	GCAAGTGGAGCCCAAGATGATCGCGGGATCGCGGCTTTCATCAAGGTGCCCAAGTACG	553		
DB 1568	GCAAGTGGAGCCCAAGATGATCGCGGGATCGCGGCTTTCATCAAGGTGCCCAAGTACG	1627		
QY 554	ACCAAGATCCTGATCGAGATCTGCGGCAAGAAGCCATCGGCACCGTGTGTGATCGGCCCA	613		
DB 1628	ACCAAGATCCTGATCGAGATCTGCGGCAAGAAGCCATCGGCACCGTGTGTGATCGGCCCA	1687		
QY 614	CCCCCGTGAATCATCGCGCGCAAAATGCTGACCCAGCTGGGCTGCAACCTGTAACTTCC	673		

QY	674	CCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATGAGCGCCCCAAGG	733
Db	1748	CCATCAGCCCATCGAGACCGTGCCGTGAAGCTGAAGCCCGCATGAGCGCCCCAAGG	1807
QY	734	TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTCGGAGGAGA	793
Db	1808	TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTCGGAGGAGA	1867
QY	794	TGGAGAGGAGGGCAAGATCACCAAGNTGGCCCCCGAGACCCCTCAACAACCCCCGTGT	853
Db	1868	TGGAGAGGAGGGCAAGATCACCAAGATCGCCCCCGAGAACCCCTCAACAACCCCCGTGT	1927
QY	854	TGCGCATCAAGAAAGGACGACCAACGATGGCGCGAAGCTGTGTGACTTCCGCGAGCTGA	913
Db	1928	TGCGCATCAAGAAAGGACGACCAACGATGGCGCGAAGCTGTGTGACTTCCGCGAGCTGA	1987
QY	914	ACAACGGACCCAGGACTTCTGGGAGTGCAGCTGGGCATCCCCACCCCGCGGCGCTGA	973
Db	1986	ACAACGGACCCAGGACTTCTGGGAGTGCAGCTGGGCATCCCCACCCCGCGGCGCTGA	2047

Qy	974	AGAAGAGAAAGAGCGTGTGACCGTGTGCGAGCGTGGCGGACGCGCTACTTCAGCGTCCCGCTGG	1033
Db	2048	AGAAGAAGAAAGAGCGTGTGACCGTGTGCGAGCGTGGCGGACGCGCTACTTCAGCGTCCCGCTGG	2107
Qy	1034	ACGAGGACITTCGCGAAGTACACCGCTTCACCAATCCCGAGCATCAACAAGAGAGACCCCG	1093
Db	2108	ACGAGGACITTCGCGAAGTACACCGCTTCACCAATCCCGAGCATCAACAAGAGAGACCCCG	2167
Qy	1094	GCATCCGCTACCAAGTACAACGCTGTGCTCCCGAGGGCTGGAAGGGCAGCCCGAGCATCTTC	1153
Db	2168	GCATCCGCTACCAAGTACAACGCTGTGCTCCCGAGGGCTGGAAGGGCAGCCCGAGCATCTTC	2227
Qy	1154	AGAGCAGCATGACCAAGATCCTGTGAGCCCTTCGCGCGCCGCAACCCGAGATCGTGATCT	1213
Db	2228	AGAGCAGCATGACCAAGATCCTGTGAGCCCTTCGCGCGCCGCAACCCGAGATCGTGATCT	2287
Qy	1214	ACCAAGTACATGGACGACCTGTACGTGGCGAGCGACCTGGAGATCGGCCAGCACCGCGCCA	1273
Db	2288	ACCA-----GGCCCCCTGTACGTGGCGAGCGACCTGGAGATCGGCCAGCACCGCGCCA	2341
Qy	1274	AGATCGAGGAGCTGCGCAAGCACACTGCTGCGCTGGGGCTTCACACCCCCGACAAGAAGC	1333
Db	2342	AGATCGAGGAGCTGCGCAAGCACACTGCTGCGCTGGGGCTTCACACCCCCGACAAGAAGC	2401
Qy	1334	ACCAGAAGGAGCCCCCTTCTCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGAACCG	1393
Db	2402	ACCAGAAGGAGCCCCCTTCTCTGCCAT-----CGAGCTGCACCCCGACAAGTGAACCG	2455
Qy	1394	TGCAGCCCATCGAGCTCGCCGAGAAAGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG	1453
Db	2456	TGCAGCCCATCGAGCTCGCCGAGAAAGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG	2515
Qy	1454	TGGCCAAAGCTGAATCTGGGCGACGAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA	1513
Db	2516	TGGCCAAAGCTGAATCTGGGCGACGAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA	2575
Qy	1514	AGCTGCTGCGCGGCGCAAGSCCTGACCGACATCTGTGCCCTTGACCGAGAGGCGCGAGC	1573
Db	2576	AGCTGCTGCGCGGCGCCAGGCCCTGACCGACATCTGTGCCCTTGACCGAGAGGCGCGAGC	2635
Qy	1574	TGGAGCTGGCGAGAAACCGGAGATCTGTGCGAGCCCGTGCAACGGCGTGTATTCAGACC	1633
Db	2636	TGGAGCTGGCGAGAAACCGGAGATCTGTGCGAGCCCGTGCAACGGCGTGTATTCAGACC	2695
Qy	1634	CCAGCAAGGACCTGCTGCGCGAGATCCAGAAAGAGGCGCACGACCACTGACGATCCAGA	1693
Db	2696	CCAGCAAGGACCTGCTGCGCGAGATCCAGAAAGAGGCGCACGACCACTGACGATCCAGA	2755
Qy	1694	TCTACCAAGGACCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGCAACCGCCC	1753
Db	2756	TCTACCAAGGACCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGCAACCGCCC	2815
Qy	1754	ACACCAAGAGCTGACGACGCTGACGAGGCGGTGCAAGAGATCGCCATCGAGAGGATCG	1813
Db	2816	ACACCAAGAGCTGACGACGCTGACGAGGCGGTGCAAGAGATCGCCATCGAGAGGATCG	2875
Qy	1814	TGATCTGGGGCAAGACCCCCAAGTTCCGCTGCCCCATCCAGAAGGAGACCTGGGAGACCT	1873
Db	2876	TGATCTGGGGCAAGACCCCCAAGTTCCGCTGCCCCATCCAGAAGGAGACCTGGGAGACCT	2935
Qy	1874	GGTGACCGCATCTCTGGCAGSCCACTTGGATCCCCGAGTGGGAGTCTGTGAAACACCCCC	1933
Db	2936	GGTGACCGCATCTCTGGCAGSCCACTTGGATCCCCGAGTGGGAGTCTGTGAAACACCCCC	2995
Qy	1934	CCCTGGTGAAGCTGTGTTACAGCTGGAGAGAGGCCATCATCGGCGCCGAGACCTTCT	1993
Db	2996	CCCTGGTGAAGCTGTGTTACAGCTGGAGAGAGGCCATCATCGGCGCCGAGACCTTCT	3053
Qy	1994	ACGTGGAAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGCGGTACTGTGACCGACC	2053
Db	3056	ACGTGGAAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGCGGTACTGTGACCGACC	3115

RESULT 13

US-10-190-435-14

03-10-190-433-14 : Sequence 14. Application US/10190435

sequence I4, Application US/10030743248A1  
: Publication No. US20030743248A1

; PUBLICATION NO: US20  
: GENERAL INFORMATION:

; GENERAL INFORMATION:

; APPLICANT: ZUR MEGEDE, Jan "

APPLICANT: BARNETT, S

; APPLICANT: LIAN, Ying

APPLICANT: ENGELBRECHT, Susan

APPLICANT: VAN RENSBURG, Estrelita J.

INVENTOR: VAN KENSBROOK, ESTHER C. ;  
: TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE

; TITLE OF INVENTION: FOLINGOCETIDES ENCOATED POLYPEPTIDES. POLYPEPTIDES. POLYPEPTIDES. POLYPEPTIDES.

FILE REFERENCE: PP18133 003 / 2302-78133

FILE REFERENCE: PP18I33.003 / 2302-18I33  
CURRENT IDENTIFICATION NUMBER: IIS/10/190 428

; CURRENT APPLICATION NUMBER

; CURRENT FILING DATE: 2002-

; NUMBER OF SEQ

; SOFTWARE: Pat

: SEO ID NO 1

LENGTH: 3537

LENGTH: 3  
TYPE: DNA

TYPE: DNA  
ORGANISM: Artificial Sequence

; ORGANISM: Ar

```
; FEATURE:
; CONVERSION OF XREFS TO SOMENAMES. GOTO WITH C
```

OTHER INFORMATION: Description of Artificially

US-10-190-435-14

Query Match 95.6%; Score 2361.2; DB 14; Length 3537;

Best Local Similarity 98.4%; Pred. No. 0;

Best local similarity	Local no. of	Mismatches	Indels	Gaps
Matches 2410: Conservative	0	28	12	2

Matches 2410; conservative 0, misclassified 20, accuracy 11)

14 TGGCCGAGGCCATGAGGCCAGGCCACAGGCCACATCCTGATGCAGCGCAGCACTTCA 73

QY 14 TGGCCGAGGCCATGAGCCAGGCCACAGGCCCAACATCCGATGCGAGCGAGCACTTCA 75

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

QY	254	AGGCCCGCGATTCCCGAGCGAGCAGAACCGCGCCCAACAGCCCAACAGCCGCGAGCTGC	313	QY	1334	ACCAGAAAGGAGCCCCCTTCTGTGTGATGGGTACGAGCTGACCCCGACAAGTGGACCG	1393
Db	1334	AGGCCCGCGAGTTCCCGAGCGAGCAGAACCGCGCCCAACAGCCCAACAGCCGCGAGCTGC	1393	Db	2408	ACCAGAAAGGAGCCCCCTTCTGTGTGATGGGTACGAGCTGACCCCGACAAGTGGACCG	2461
QY	314	AGGTGCGCGCGCAACACCCCGCAGAGAGCGCGCGCGAGCCAGCGGCAACCTGAAT	373	QY	1394	TGCAGCCCATCGAGCTGCCGGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG	1453
Db	1394	AGGTGCGCGCGCAACACCCCGCAGAGAGCGCGCGCGAGCCAGCGGCAACCTGAAT	1453	Db	2462	TGCAGCCCATCGAGCTGCCGGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG	2521
QY	374	TCCCCCAGATCACCTGTGAGCGCCCTGTGTGAGCATCAAGGTGGGCGGCGAGATCA	433	QY	1454	TGGCAAGCTGAATGGGCGCAGCGAGATCAACCCCGCATCAAGGTGGGCGGCGAGTGTGCA	1513
Db	1454	TCCCCCAGATCACCTGTGAGCGCCCTGTGTGAGCATCAAGGTGGGCGGCGAGATCA	1513	Db	2522	TGGCAAGCTGAATGGGCGCAGCGAGATCAACCCCGCATCAAGGTGGGCGGCGAGTGTGCA	2581
QY	434	AGGAGGCTGTCTGGAACAACCGCGCGCGAGCAACACCTGTGTGAGGAGATGAGCTCCCG	493	QY	1514	AGCTGTGCGGCGCGCAGAGCCCTGACCGACATCTGTCCTTACCGAGAGGCGCGAGC	1573
Db	1514	AGGAGGCTGTCTGGAACAACCGCGCGCGAGCAACACCTGTGTGAGGAGATGAGCTCCCG	493	Db	2582	AGCTGTGCGGCGCGCAGAGCCCTGACCGACATCTGTCCTTACCGAGAGGCGCGAGC	2641
QY	494	GCAAGTGGAGCCCAAGATGATCGCGGCGCATCGCGGCTTCAACAGTGGCGCATAGC	553	QY	1574	TGGAGCTGGCGAGAACCGCGAGATCTTGGCGAGCCGTGACCGCGTGTACTACGACC	1633
Db	1574	GCAAGTGGAGCCCAAGATGATCGCGGCGCATCGCGGCTTCAACAGTGGCGCATAGC	1633	Db	2642	TGGAGCTGGCGAGAACCGCGAGATCTTGGCGAGCCGTGACCGCGTGTACTACGACC	2701
QY	554	ACCAGATCCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTCTGATCGGCCCA	613	QY	1634	CCAGCAAGGACTGTGTGGCGGAGATCCAGAGCAGGCGCCACGACAGTGGACCTTACCAGA	1693
Db	1634	ACCAGATCCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTCTGATCGGCCCA	613	Db	2702	CCAGCAAGGACTGTGTGGCGGAGATCCAGAGCAGGCGCCACGACAGTGGACCTTACCAGA	2761
QY	614	CCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGTGACACCTGAATTC	673	QY	1694	TCTACAGAGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACGCCCAAGTGCACCCGCC	1753
Db	1694	CCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGTGACACCTGAATTC	1753	Db	2762	TCTACAGAGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACGCCCAAGTGCACCCGCC	2821
QY	674	CCATCAGCCCCATCGAGACCGTGCCTGAGCTGGAAGCCCGGCATGAGCGGCCCAAGG	733	QY	1754	ACACCAACGAGCGTGAAGCAGCTGACCGAGGCGGTGAGAGATCGCCATGGAGAGCATCG	1813
Db	1754	CCATCAGCCCCATCGAGACCGTGCCTGAGCTGGAAGCCCGGCATGAGCGGCCCAAGG	1813	Db	2822	ACACCAACGAGCGTGAAGCAGCTGACCGAGGCGGTGAGAGATCGCCATGGAGAGCATCG	2881
QY	734	TGAAGAGTGGCCCTGACCGAGAGAGATCAAGGCGCTGACCGCCATCTGCGAGGAGA	793	QY	1814	TGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCT	1873
Db	1814	TGAAGAGTGGCCCTGACCGAGAGAGATCAAGGCGCTGACCGCCATCTGCGAGGAGA	1873	Db	2882	TGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCT	2941
QY	794	TGGAGAGGAGGCAAGATCAACAGATCGGCGCCCGAGAACCCCTACAAACCCCGCTGT	853	QY	1874	GGTGAACCGACTACTCTGGCAGGCGCCCTGATTCGCGAGTGGGAGTTGTTGAACACCCCG	1933
Db	1874	TGGAGAGGAGGCAAGATCAACAGATCGGCGCCCGAGAACCCCTACAAACCCCGCTGT	1933	Db	2942	GGTGAACCGACTACTCTGGCAGGCGCCCTGATTCGCGAGTGGGAGTTGTTGAACACCCCG	3001
QY	854	TCGCCATCAAGAGAGAGACACCAAGTGGCGCGAGCTGGGAGCTTCCGCGAGCTGA	913	QY	1934	CCCTGGTGAAGCTGTGTACAGCTGGAGAGAGGAGCCCATCATCGGCGCGAGAGCTTCT	1993
Db	1934	TCGCCATCAAGAGAGAGACACCAAGTGGCGCGAGCTGGGAGCTTCCGCGAGCTGA	1993	Db	3002	CCCTGGTGAAGCTGTGTACAGCTGGAGAGAGGAGCCCATCATCGGCGCGAGAGCTTCT	3061
QY	914	ACAAGCGCACCCAGACTTCTGGAGGTGACAGCTGGGCTATCCCGACCCCGCGGCTGA	973	QY	1994	ACGTGGAAGGCGCCCGCCAAACCGCGAGACCAAGATCGGCAAGGCGCGCTACGTGACCGG	2053
Db	1994	ACAAGCGCACCCAGACTTCTGGAGGTGACAGCTGGGCTATCCCGACCCCGCGGCTGA	2053	Db	3062	ACGTGGAAGGCGCCCGCCAAACCGCGAGACCAAGATCGGCAAGGCGCGCTACGTGACCGG	3121
QY	974	AGAAGAAGAGCGTGACCGTGTGAGCGTGGGCGACGCTTCTCAGCGTGGCCCTGG	1033	QY	2054	GGGGCGGCGAGAGATCGTGAAGCTGACCGAGACCCCAACCAAGAGAGCCGAGTGCAGG	2113
Db	2054	AGAAGAAGAGCGTGACCGTGTGAGCGTGGGCGACGCTTCTCAGCGTGGCCCTGG	2113	Db	3122	GGGGCGGCGAGAGATCGTGAAGCTGACCGAGACCCCAACCAAGAGAGCCGAGTGCAGG	3181
QY	1034	ACGAGGATTCGCGAAGTACACCGCTTCAACATCCCGAGCATCAACAGAGAGCCCG	1093	QY	2114	CCATCCAGCTGGCCCTGAGGACAGCGGACAGCGAGTGAACATCTGTGACCGAGAGCTTCT	2173
Db	2114	ACGAGGATTCGCGAAGTACACCGCTTCAACATCCCGAGCATCAACAGAGAGCCCG	2173	Db	3182	CCATCCAGCTGGCCCTGAGGACAGCGGACAGCGAGTGAACATCTGTGACCGAGAGCTTCT	3241
QY	1094	GCATCCGCTACAGTACAACTGTGTGCGCCCGAGGCTGGAAGGCGAGCCCGAGCATCTTC	1153	QY	2174	AGCCCTGGGCGATCATCCAGGCGCAGCGGACAGAGCGAGCGAGCTGTGTGAACAGG	2233
Db	2174	GCATCCGCTACAGTACAACTGTGTGCGCCCGAGGCTGGAAGGCGAGCCCGAGCATCTTC	2233	Db	3242	AGCCCTGGGCGATCATCCAGGCGCAGCGGACAGAGCGAGCGAGCTGTGTGAACAGG	3301
QY	1154	AGAGCAGCATGACCAAGATCTGTGAGCGCTTCCGCGCGCGCAACCGCGAGTCTGTCT	1213	QY	2234	TCATCAGCAGCTGTATCAAGAGGAGAGGTGTACTTGTGCTGGTGGTGGTGGTGGTGGTGG	2293
Db	2234	AGAGCAGCATGACCAAGATCTGTGAGCGCTTCCGCGCGCGCAACCGCGAGTCTGTCT	2293	Db	3302	TCATCAGCAGCTGTATCAAGAGGAGAGGTGTACTTGTGCTGGTGGTGGTGGTGGTGGTGG	3361
QY	1214	ACCAGTACATGAGCAGCTGTGTGAGCGAGCTGAGATCGGCGAGCAGCGCGCA	1273	QY	2294	GCATCGGCGGCAACGAGCAGATCGAAGCTGGTGAAGCAAGGCGATCGCAGAGTGTGT	2353
Db	2294	ACCAGTACATGAGCAGCTGTGTGAGCGAGCTGAGATCGGCGAGCAGCGCGCA	2347	Db	3362	GCATCGGCGGCAACGAGCAGATCGAAGCTGGTGAAGCAAGGCGATCGCAGAGTGTGT	3421
QY	1274	AGATCGAGAGTGTGCGAAGCACTGTGTGCGTGGGGCTTCCACCCCGCGAGAGAGC	1333	QY	2354	TCTGAGAGCGCATCGATGGCGCATCTGATCTACAGTACATGAGCAGACCTGTACGTGG	2413
Db	2348	AGATCGAGAGTGTGCGAAGCACTGTGTGCGTGGGGCTTCCACCCCGCGAGAGAGC	2407	Db	3422	TCTGAGAGCGCATCGATGGCGCATCTGATCTACAGTACATGAGCAGACCTGTACGTGG	3481
				QY	2414	GCAGCGGCGGCGCTTAGGATCGATTAAAGCTTCCCGGGGCGCTAGACCCGT	2463

Db	3482	GCAGCGCGCCCTAGGATCATTAAGCTTCCGGGGGTAGCACGGT	3531
RESULT 14			
US-10-190-435-15			
Sequence 15, Application US/10190435			
Publication No. US20030143248A1			
GENERAL INFORMATION:			
APPLICANT: ZUR MEGEDE, Jan			
APPLICANT: BARNETT, Susan W.			
APPLICANT: LIAN, Ying			
APPLICANT: ENGELBRECHT, Susan			
APPLICANT: VAN RENSBURG, Estrelita J.			
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C			
TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF			
FILE REFERENCE: PP18133.003 / 2302-18133			
CURRENT APPLICATION NUMBER: US/10/190,435			
CURRENT FILING DATE: 2002-12-30			
NUMBER OF SEQ ID NOS: 319			
SOFTWARE: PatentIn Ver. 2.0			
SEQ ID NO 15			
LENGTH: 3537			
TYPE: DNA			
ORGANISM: Artificial Sequence			
FEATURE:			
OTHER INFORMATION: Description of Artificial Sequence: GagPolmutIna_C			
US-10-190-435-15			
Query Match		95.6%;	Score 2361.2; DB 14; Length 3537;
Best Local Similarity		98.4%;	Pred. No. 0;
Matches 2410; Conservative 0; Mismatches 28; Indels 12; Gaps 3;			
QY	14	TGGCCGAGGCGCATGAGCCAGCCACAGCCACCATCTGTGATGCGAGGGGCAACTTCA	73
Db	1094	TGGCCGAGGCGATAGCCAGCCACAGCCACCATCTGTGATGCGAGGGGCAACTTAAA	1153
QY	74	AGGCCCCCAAGCGCATCATCAAGTGTCTTAACTCGCGCAAGGAGGGGCACATCGCCGCA	133
Db	1154	AGGCCCCCAAGCGCATCATCAAGTGTCTTAACTCGCGCAAGGAGGGGCACATCGCCGCA	1213
QY	134	ACTGCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGGCAAGGAGGGGCACCATGATGA	193
Db	1214	ACTGCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGGCAAGGAGGGGCACCATGATGA	1273
QY	194	AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCCGAGGACCTTGGCTTCCCCCAGAGCCA	253
Db	1274	AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCCGAGGACCTTGGCTTCCCCCAGAGCCA	1333
QY	254	AGGCCGCGAGTTCCCGAGGAGCAGAAACCGCGCAACAGCCCCCAACAGCGCGAGCTGC	313
Db	1334	AGGCCGCGAGTTCCCGAGGAGCAGAAACCGCGCAACAGCCCCCAACAGCGCGAGCTGC	1393
QY	314	AGGTGCGCGGGCAAAACCCCGAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	373
Db	1394	AGGTGCGCGGGCAAAACCCCGAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1453
QY	374	TCCCCCAGATCACCTGTGGCAGCGCCCCCTTGTGTGATCATCAAGGTGGCGGGCCAGATCA	433
Db	1454	TCCCCCAGATCACCTGTGGCAGCGCCCCCTTGTGTGATCATCAAGGTGGCGGGCCAGATCA	1513
QY	434	AGGAGGCCCTGCTGGAACACCGCGCGCGCAACGACCGTGTGGAGAGATGAGCCCTGCCCG	493
Db	1514	AGGAGGCCCTGCTGGAACACCGCGCGCGCAACGACCGTGTGGAGAGATGAGCCCTGCCCG	1573
QY	494	GCAAGTGGAAAGCCCAAGATGATGCGCGGCATCGCGGGCTTCATCAAGGTGGCGGCAGTACG	553
Db	1574	GCAAGTGGAAAGCCCAAGATGATGCGCGGCATCGCGGGCTTCATCAAGGTGGCGGCAGTACG	1633
QY	554	ACCGATCTCTGATCGAGATCTTGGCGCAAGAGAGCCATCGGCACCGTGTGATCGGCCCA	613
Db	1634	ACCGATCTCTGATCGAGATCTTGGCGCAAGAGAGCCATCGGCACCGTGTGATCGGCCCA	1693

```
Db 2762 TCTACGAGGACCTTCAAGAACCTGAGAACCGGCAAGTAGTACGCCAAGATCGCACCGCC 2821
Qy 1754 ACACCAACGAGCTGAAGCAGCTGACCGAGGCGGTGCAAGAGATGCCCATCGAGAGCATCG 1813
Db 2822 ACACCAACGAGCTGAAGCAGCTGACCGAGGCGGTGCAAGAGATGCCCATCGAGAGCATCG 2881
Qy 1814 TGATCTGGGGCAAGACCCCCCAAGTTCCGCTGCCCCATCCAGAGGAGACCTGGGAGACCT 1873
Db 2882 TGATCTGGGGCAAGACCCCCCAAGTTCCGCTGCCCCATCCAGAGGAGACCTGGGAGACCT 2941
Qy 1874 GGTGACCGACTACTGGCAGGCCACCTGATGCCCGGAGTGGAGTTGCTGAACACCCCC 1933
Db 2942 GGTGACCGACTACTGGCAGGCCACCTGATGCCCGGAGTGGAGTTGCTGAACACCCCC 3001
Qy 1934 CCCTGGTGAAGCTGTGTATCCAGCTGGAGAGAGGCCCATCATCGGCCCGAGACCTTCT 1993
Db 3002 CCCTGGTGAAGCTGTGTATCCAGCTGGAGAGAGGCCCATCATCGGCCCGAGACCTTCT 3061
Qy 1994 ACCTGGAGCGCGCGCCCAACCGCAGACCAAGATCGGCNAGCGCGGTAGTGACCGACC 2053
Db 3062 ACCTGGAGCGCGCGCCCAACCGCAGACCAAGATCGGCNAGCGCGGTAGTGACCGACC 3121
Qy 2054 GGGGCGCGCAGAAAGTCTGTAGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGG 2113
Db 3122 GGGGCGCGCAGAAAGTCTGTAGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGG 3181
Qy 2114 CATTCAGCTGGCTTCAGAGACGCGCAGCAGAGGTGAACATCTGACCGACAGCCAGT 2173
Db 3182 CCATTCAGCTGGCTTCAGAGACGCGCAGCAGAGGTGAACATCTGACCGACAGCCAGT 3241
Qy 2174 ACGCCCTGGGATCATCCAGCGCCAGCCGACCAAGAGCGAGCGAGCTGTGAACAGGA 2233
Db 3242 ACGCCCTGGGATCATCCAGCGCCAGCCGACCAAGAGCGAGCGAGCTGTGAACAGGA 3301
Qy 2234 TCATCGACGCTGATCAAGAGAGAGAGAGGTGTACTGAGTGGTGGTGGCCGCCCAAGG 2293
Db 3302 TCATCGACGCTGATCAAGAGAGAGAGGTGTACTGAGTGGTGGTGGCCGCCCAAGG 3361
Qy 2294 GCATCGCGCGCAACGAGCAGATCGAAGCTGGTGAAGAGGATCGCGCAAGGTGCTGT 2353
Db 3362 GCATCGCGCGCAACGAGCAGATCGAAGCTGGTGAAGAGGATCGCGCAAGGTGCTGT 3421
Qy 2354 TCCTCGACGCTGATCGCGGCTGATGATCTACAGTACAGTACATGAGACCTGTACGTGG 2413
Db 3422 TCCTCGACGCTGATCGCGGCTGATGATCTACAGTACAGTACATGAGACCTGTACGTGG 3481
Qy 2414 GCAGCGCGCGCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2463
Db 3482 GCAGCGCGCGCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 3531
```

RESULT 15

```
US-10-190-435-12
; Sequence 12, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEYER, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Artificial Sequence
```

```
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GagCompPolmutInaRevNef_C
US-10-190-435-12
```

```
Query Match 95.6%; Score 2360.2; DB 14; Length 5145;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2397; Conservative 0; Mismatches 8; Indels 12; Gaps 2;

Qy 14 TGGCCGAGCCATGAGCCAGCCACAGCGCAACATCTCTGATGAGCGAGCAACTTCA 73
Db 1487 TGGCCGAGCCATGAGCCAGCCACAGCGCAACATCTCTGATGAGCGAGCAACTTCA 1546
Qy 74 AGGGCCCCAAGCGCATCATCAAGTCTTCAACTCGGCAAGAGGCGCACATCGCCCGCA 133
Db 1547 AGGGCCCCAAGCGCATCATCAAGTCTTCAACTCGGCAAGAGGCGCACATCGCCCGCA 1606
Qy 134 ACTGCCGCGCCCCCGCAAGAGGCTCTGGAAGTGGCGCAAGAGGCGCACAGATGA 193
Db 1607 ACTGCCGCGCCCCCGCAAGAGGCTCTGGAAGTGGCGCAAGAGGCGCACAGATGA 1666
Qy 194 AGGACTGACCGAGCGCGCAGGCAACTTCTCCGCGAGGACCTGGCTTCCCGCAGGCA 253
Db 1667 AGGACTGACCGAGCGCGCAGGCAACTTCTCCGCGAGGACCTGGCTTCCCGCAGGCA 1726
Qy 254 AGGCCCGGAGTTCCCGCAGCGAGCAAGAACCGCGCAACAGCCCCACAGCGCGAGCTGC 313
Db 1727 AGGCCCGGAGTTCCCGCAGCGAGCAAGAACCGCGCAACAGCCCCACAGCGCGAGCTGC 1786
Qy 314 AGGTGGCGGCGCAACACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373
Db 1787 AGGTGGCGGCGCAACACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846
Qy 374 TCCCCCAGATCACCTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433
Db 1847 TCCCCCAGATCACCTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1906
Qy 434 AGGAGCCCTGTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493
Db 1907 AGGAGCCCTGTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1966
Qy 494 GGAAGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 553
Db 1967 GGAAGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2026
Qy 554 ACCAGATCTGTGATGAGATCTCGCGCAAGAGGCGCATCGGACCGCTGTGTATCGCGCCCA 613
Db 2027 ACCAGATCTGTGATGAGATCTCGCGCAAGAGGCGCATCGGACCGCTGTGTATCGCGCCCA 2086
Qy 614 CCCCCGTGAACATCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 673
Db 2087 CCCCCGTGAACATCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2146
Qy 674 CCATCAGCCCCCATCGAGACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 733
Db 2147 CCATCAGCCCCCATCGAGACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2206
Qy 734 TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 793
Db 2207 TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 2266
Qy 794 TGGAGAAGGAGGCGCAAGATCACCAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 853
Db 2267 TGGAGAAGGAGGCGCAAGATCACCAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2326
Qy 854 TCGCCCATCAAGAGAGAGGAGCAGACCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 913
Db 2327 TCGCCCATCAAGAGAGAGGAGCAGACCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2386
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGAGTCCCGCGCGCGCGCGCGCGCGCG 973
Db 2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGAGTCCCGCGCGCGCGCGCGCGCGCG 2446
```

QY 974 AGAAGAAAGAGCGTGTACCGTGTCTGGACGTGGGCGACGCTACTTTCAGCGTGGCCCTGG 1033  
Db 2447 AGAAGAAAGAGCGTGTACCGTGTCTGGACGTGGGCGACGCTACTTTCAGCGTGGCCCTGG 2506  
QY 1034 ACGAGGACTTCGCAAGTACACCGCTTTCACCATCCCCAGCATCAACAAACAGACCCCGG 1093  
Db 2507 ACGAGGACTTCGCAAGTACACCGCTTTCACCATCCCCAGCATCAACAAACAGACCCCGG 2566  
QY 1094 GCATCCGCTACAGTACAACTGTCTGCCCGCCAGGGGTGAAGGGAGCGCCCGAGCATCTTCC 1153  
Db 2567 GCATCCGCTACAGTACAACTGTCTGCCCGCCAGGGGTGAAGGGAGCGCCCGAGCATCTTCC 2626  
QY 1154 AGAGCAGCATGACCAAGATCTCTGAGCGCTTCCGCGCCCGCAACCCCGAGATCTGTATCT 1213  
Db 2627 AGAGCAGCATGACCAAGATCTCTGAGCGCTTCCGCGCCCGCAACCCCGAGATCTGTATCT 2686  
QY 1214 ACCAGTACATGAACGACTGTACGTGGGCGACGACCTTGGAGATCGGCAGACACCGCGCCA 1273  
Db 2687 ACCA-----GGCCCCCTGTACGTGGGCGACGACCTTGGAGATCGGCAGACACCGCGCCA 2740  
QY 1274 AGATCGAGGAGCTGCGCAAGCACCTCTGCGCGCCGCTTCCGCGCCCGCAACCCCGAGAGC 1333  
Db 2741 AGATCGAGGAGCTGCGCAAGCACCTCTGCGCGCCGCTTCCGCGCCCGCAACCCCGAGAGC 2800  
QY 1334 ACCAGAAGAGCCCCCTTCTCTGTGGATGGGCTACGAGCTGCACCCCGACAAAGTGGACCG 1393  
Db 2801 ACCAGAAGAGCCCCCTTCTCTGTGGAT-----CGAGCTGCACCCCGACAAAGTGGACCG 2854  
QY 1394 TCCAGCCCATCGAGCTGCCGAGAGAGAGCTGACCGTGAACGACATCCAGAGCTGG 1453  
Db 2855 TGCAGCCCATCGAGCTGCCGAGAGAGAGCTGACCGTGAACGACATCCAGAGCTGG 2914  
QY 1454 TGGGCAAGCTGAATGGGCGACAGATCTACCCCGGCATCAAGGTGGCGCGAGCTGTGCA 1513  
Db 2915 TGGGCAAGCTGAATGGGCGACAGATCTACCCCGGCATCAAGGTGGCGCGAGCTGTGCA 2974  
QY 1514 AGCTGCTGGCGGCGCGCAAGGCCCTGACCGACATCGTGCCTTGAACGAGAGGCCGAGC 1573  
Db 2975 AGCTGCTGGCGGCGCGCAAGGCCCTGACCGACATCGTGCCTTGAACGAGAGGCCGAGC 3034  
QY 1574 TGGAGTGGCCGAGAACCGCGAGATCTTCCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633  
Db 3035 TGGAGTGGCCGAGAACCGCGAGATCTTCCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094  
QY 1634 CCAGCAAGGACCTGTGTGGCCGAGATCCAGAGCAGGGCCACGACAGTGGACCTTACCAGA 1693  
Db 3095 CCAGCAAGGACCTGTGTGGCCGAGATCCAGAGCAGGGCCACGACAGTGGACCTTACCAGA 3154  
QY 1694 TCTACAGGAGCCCTTCAAGACCTGAAGACCGGCAAGTACGCCAAGATGGCACCGGCC 1753  
Db 3155 TCTACAGGAGCCCTTCAAGACCTGAAGACCGGCAAGTACGCCAAGATGGCACCGGCC 3214  
QY 1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCGGTGCAGAAATCGCCATGGAGAGCATCG 1813  
Db 3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCGGTGCAGAAATCGCCATGGAGAGCATCG 3274  
QY 1814 TGATCTGGGCGAAGACCCCGAGTTCGCCCTGCCATCCAGAGGAGACCTGGGAGCCT 1873  
Db 3275 TGATCTGGGCGAAGACCCCGAGTTCGCCCTGCCATCCAGAGGAGACCTGGGAGCCT 3334  
QY 1874 GGTGGAACCGACTACTTGGCAGGCGACCTGTGATCCCCGAGTGGGAGTTCTGTAACCCCCC 1933  
Db 3335 GGTGGAACCGACTACTTGGCAGGCGACCTGTGATCCCCGAGTGGGAGTTCTGTAACCCCCC 3394  
QY 1934 CCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGGCCATCATCGGCGCCGAGACCTTCT 1993  
Db 3395 CCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGGCCATCATCGGCGCCGAGACCTTCT 3454  
QY 1994 ACGTGAACCGGCGCGCCAAACCGCGAGACCAAGATCGGAAAGCCGGCTACGTGACCGACC 2053  
Db 3455 ACGTGAACCGGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514  
QY 2054 GGGGCGCGCAGAAAGATCGTGAAGCCTTGAACGAGACCAACCAACAGAAAGCCGAGCTGACG 2113

Db 3515 GGGGCGCGCAGAAAGATCGTGAAGCCTTGAACGAGACCAACCAAGAACCGAGCTGACG 3574  
QY 2114 CCATCCAGCTGGCCCTTGCAGGACAGCGGCGAGCGGTGAACATCGTGACCGACAGCCAGT 2173  
Db 3575 CCATCCAGCTGGCCCTTGCAGGACAGCGGCGAGCGGTGAACATCGTGACCGACAGCCAGT 3634  
QY 2174 ACGCCCTGGGCATCATCCAGGCGCCGACCAAGAGCGAGCGAGCTGGTGAACCCAGA 2233  
Db 3635 ACGCCCTGGGCATCATCCAGGCGCCGACCAAGAGCGAGCGAGCTGGTGAACCCAGA 3694  
QY 2234 TCATCCAGCAGCTGTATCAAGAAAGGAGAGGTGTACTCTGAGCTGGGTGCCGCCCAAGG 2293  
Db 3695 TCATCCAGCAGCTGTATCAAGAAAGGAGAGGTGTACTCTGAGCTGGGTGCCGCCCAAGG 3754  
QY 2294 GCATCCGCGCAACGAGCAGATCGACAGCTGGTGACAAAGGCGCATCCGCAAGGTGCTGT 2353  
Db 3755 GCATCCGCGCAACGAGCAGATCGACAGCTGGTGACAAAGGCGCATCCGCAAGGTGCTGT 3814  
QY 2354 TCCTGGACCGGCATCGATGGCGGCATCGTGATCTACCAAGTACATGGACGACCTGTACG 2413  
Db 3815 TCCTGGACCGGCATCGATGGCGGCATCGTGATCTACCAAGTACATGGACGACCTGTACG 3874  
QY 2414 GCAGCGCGGCCCTTAGG 2430  
Db 3875 GCAGCGCGGCCCTTAGG 3891

Search completed: April 10, 2004, 21:24:32  
Job time : 604.779 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 06:34:42 ; Search time 4127.92 Seconds  
(without alignments)  
17774.420 Million cell updates/sec

Title: US-09-610-313-32

Perfect score: 2457  
Sequence: 1 gtgacgccaccatggccga.....gggtcgtaccggtgaattc 2457

Scoring table:

IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estm:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_estc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_est3:\*

12: gb\_est4:\*

13: gb\_est5:\*

14: gb\_est6:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	90.2	3.7	2598	11 AY103647	AY103647 Zea mays
2	85.6	3.5	951	12 BM321451	BM321451 rockefell
3	80.6	3.3	869	14 CK159167	CK159167 FGAS04056
4	80.2	3.3	1132	12 BM320864	BM320864 rockefell

5	79.6	3.2	1165	12 BM320900	BM320900 rockefell
6	79	3.2	867	12 BM321430	BM321430 rockefell
7	76.6	3.1	1550	12 BM321022	BM321022 rockefell
8	75	3.1	545	12 BT724851	BT724851 1031075E0
9	75	3.1	862	12 BM321023	BM321023 rockefell
10	74.4	3.0	914	28 BZ568300	BZ568300 Pac2-164
11	73.8	3.0	2433	29 AY401196	AY401196 Hmc sapi
12	73.2	3.0	853	12 BM321393	BM321393 rockefell
13	72.8	3.0	566	12 BM587428	BM587428 170006873
14	72.6	3.0	788	14 CB643171	CB643171 OSJNEB03L
15	72.4	3.0	753	29 CC675888	CC675888 OGWCO51TH
16	72.2	2.9	640	10 BE601575	BE601575 HVSME009
17	72.2	2.9	688	14 CB648640	CB648640 OSJNEB12C
18	72.2	2.9	764	14 CB651670	CB651670 OSJNEB16L
19	72.2	2.9	766	14 CB642928	CB642928 OSJNEB03F
20	72.2	2.9	809	14 CB641397	CB641397 OSJNEB01A
21	72.2	2.9	841	14 CB651502	CB651502 OSJNEB16H
22	72.2	2.9	851	29 CG260623	CG260623 OGWGE95TV
23	71.8	2.9	1304	28 BZ577729	BZ577729 msh2 5546
24	71.8	2.9	1333	11 AY325173	AY325173 Rattus no
25	71.6	2.9	500	12 BM372120	BM372120 EBR003_SQ
26	71.6	2.9	538	12 BM368580	BM368580 EBR008_SQ
27	71.6	2.9	540	9 AJ471121	AJ471121
28	71.6	2.9	566	13 BQ464692	BQ464692 HF02P20r
29	71.6	2.9	575	13 BU976068	BU976068 HA03B08r
30	71.6	2.9	576	13 BU984666	BU984666 HF04K03r
31	71.6	2.9	578	12 BM377112	BM377112 EBR005_SQ
32	71.6	2.9	579	10 BF253640	BF253640 HVSME1000
33	71.6	2.9	582	12 BM372871	BM372871 EBR004_SQ
34	71.6	2.9	586	13 BU994273	BU994273 HM06H03r
35	71.6	2.9	588	13 BQ765498	BQ765498 EBR003_SQ
36	71.6	2.9	596	13 BU997745	BU997745 HI08O07r
37	71.6	2.9	600	9 AL508778	AL508778 AL508778
38	71.6	2.9	615	13 CA009718	CA009718 HU14O18r
39	71.6	2.9	621	13 BU986998	BU986998 HF13I08r
40	71.6	2.9	625	13 BQ768295	BQ768295 EBR008_SQ
41	71.6	2.9	637	13 BU999333	BU999333 HI14D01r
42	71.6	2.9	641	13 BU978712	BU978712 HA14B02r
43	71.6	2.9	645	13 BU977527	BU977527 HA11J06r
44	71.6	2.9	646	13 BU968956	BU968956 HB10B12r
45	71.6	2.9	655	13 BU998444	BU998444 HB13K04r

# ALIGNMENTS

RESULT 1	AY103647	AY103647	2598 bp	mRNA	linear	HTC 16-OCT-2002
LOCUS	Zea mays	PCO142084	mRNA sequence.			
DEFINITION	Zea mays	PCO142084	mRNA sequence.			
ACCESSION	AY103647					
VERSION	AY103647.1	GI:21206725				
KEYWORDS	HTC.					
SOURCE	Zea mays					
ORGANISM	Zea mays					
REFERENCE	1 (bases 1 to 2598)					
AUTHORS	Hainey C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.					
TITLE	Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes					
JOURNAL	Unpublished (2002)					
REFERENCE	2 (bases 1 to 2598)					
AUTHORS	Coe, E.H.					
TITLE	Direct Submission					
JOURNAL	Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA					
COMMENT	If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the					



QY 2150 CCGACAGCAGTACGCCCTGGGATCATCCAGGCCCGCCGACAGAGCGAGCGAGC 2209  
Db 1902 CTTCGGTGTCTCAAGATCACCAGGTTCCAGGAGGAGCTCCCGCGGTGCTGCCCCAGG 1961  
QY 2210 TGGTGAA 2216  
Db 1962 AGGTGA 1968

RESULT 2  
BM321451  
LOCUS  
DEFINITION  
Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC 3.3.1.1), mRNA sequence.  
BM321451  
BM321451.1 GI:18055857  
EST.  
Mastigamoeba balamuthi  
Mastigamoeba balamuthi  
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.  
REFERENCE  
1 (bases 1 to 951)  
Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Durufle, L., Gaasterland, J., Lopez, P., Muller, M. and Philippe, H.  
The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba  
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
21819461  
11830664  
Contact: Muller Miklos  
Laboratory of Biochemical Parasitology  
The Rockefeller University  
1230 York Avenue, New York, NY 10021, USA  
Email: mmuller@rockvax.rockefeller.edu  
Insert Length: 951 Std Error: 0.00  
POLYA-No. Location/Qualifiers  
1..951  
/organism="Mastigamoeba balamuthi"  
/mol\_type="mRNA"  
/strain="ATCC 30984"  
/db\_xref="taxon:108607"  
/clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
/note="syn: Phreatamoeba balamuthi"

FEATURES  
source  
1..951  
Query Match 3.5%; Score 85.6; DB 12; Length 951;  
Best Local Similarity 46.0%; Pred. No. 0.69;  
Matches 323; Conservative 0; Mismatches 376; Indels 3; Gaps 1;

QY 1722 CGCCAGATCGGACCGCCCGACACCGACGAGCGTGAAGCGGCGGTGCAGAA 1781  
Db 219 CGCCAGATCGGTCGTGGTGTCTGTCGACATCTTCTCGACGAGCACCACGCGCGCGC 278  
QY 1782 GATCGCATGAGAGATCTGATCTGGGGCAGACCCCAAGTTCCGCTGCCATCA 1841  
Db 279 CATCGCGCAGCGCGGTCTCGGTCTCGCTGGAAGGGCGAGAACTCCAGAGTACTG 338  
QY 1842 GAAGGAGACCTGGGAGACCTCTGTCGACCGCACTACTGCGAGCGCCACCTGGATCCCCGAGTG 1901  
Db 339 GGAGTGACCTGGAAGCGCCCTGTCTCGCCCTTACCAGGGCCCTCAGATCATCTGCA 398  
QY 1902 GGAGTTGTAACACCCCGCCCTCGTGAAGCTGTGTATCCAGCTGGAGAGAGCGCCAT 1961  
Db 399 CGACGGCGGTGACCGCACTCTGATGATCCCAAGGGGTTTCGCGCGCGAGCAACCCCAA 458  
QY 1962 CATCGGGCGCGAGACCTTCTACGTGACAGCGCGCCCAACCGGAGACCAAGATCGCAA 2021  
Db 459 GCTGCTGGAGACGACAGAGGGCTCGAGGAGTCTGCTGCTCAACAACTGCTCAAGCA 518  
QY 2022 GGCGGGCTACTGTACCGACCGGGGCGCGCAGAGATCGTGAGCTTACCGAGACCA 2081

Db 519 GGTCCAGAGGAGAGCGCCGCTTCTGGCACAAGATCCTCCCGAGATCCGCGGTGTGAG 578  
QY 2082 CCAGAGAGCGAGCTGAGCGCATCCAGCTGGCCCTGTCAGAGACAGCGGAGCGAGTGAA 2141  
Db 579 CGAGGAGACGACGACTGGCGTATGAGGCTGTACCACTGTCACCGGACGCGCAAGCTGCT 638  
QY 2142 CATGTGACGAGACGACGAGTACGCTGGGCTCATCCAGCGCCCGACGCGCCAGAGCGGA 2201  
Db 639 GTTCCCGCGCTCAACGTCACGACGAC---TCTNTCACCAGAGCAAGTTCCACACATCTA 695  
QY 2202 GAGCGAGCTGGTGAACAGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTTACCTGAG 2261  
Db 696 CGGCTGCGCGCCTCATCGCTGATCCAGCGCATCAAGCGCGGACCGACGCTGATGCTCGCGCG 755  
QY 2262 CTGAGTGGCGCGCCCAAGAGGATCGCGGCAAGAGCAGATCGACAAGCTGGTGAGCAA 2321  
Db 756 CAAGTGTCCGCTGCTCGCGGCTACGCGGAGCTGGGCAAGGGCTGCGCCGAGTCTGCTGCG 815  
QY 2322 GGGCATCGGCAAGTGTCTTCTCGGACGCGATCGATGCGGCGCATCTGATACCAAGTA 2381  
Db 816 CGGCGAGGGCTGCGCGCTCATCTGTCAGCGAGATCGACCCATCTGCGCGCTGCAGCGGCTC 875  
QY 2382 CATGAGCAGCTGTACCTGGGCGCGCGCGCTTAGGATCGA 2423  
Db 876 GATGCGCGGCTTCGAGGTCAACGCTCGAGCGCGGGGCTCGA 917

RESULT 3  
CK159167  
LOCUS  
DEFINITION  
FGAS040564 Triticum aestivum FGAS: Talt5 Triticum aestivum CDNA,  
mRNA sequence.  
CK159167  
CK159167.1 GI:38985053  
EST.  
Triticum aestivum (bread wheat)  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooideae; Triticeae; Triticum.  
1 (bases 1 to 869)  
Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Larocque, A.,  
Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,  
Penniket, C., Roach, J.L. and Sarhan, F.  
Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
Unpublished (2003)  
Contact: Wm L Crosby  
Bioinformatics  
University of Saskatchewan, Department of Computer Science  
1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769  
Fax: 306 966 2033  
Email: fgas\_est@cs.usask.ca  
This sequence is the direct result of the Base calling software  
Phred (default parameters). It is the raw base calls. To aid in the  
identification of the high quality insert the software Lucy  
(default parameters) has been run on this sequence. Lucy identified  
the region (128,636).  
Plate: Talt537 row: N column: 23.  
Location/Qualifiers  
1..869  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Wheat line PI 178383"  
/db\_xref="taxon:4565"  
/lab\_host="DHS alpha"  
/clone\_lib="Triticum aestivum FGAS: Talt5"  
/note="Organ: Crown; Vector: pGEM-T; SSH (suppression  
subtractive hybridization) cDNA library from genotype  
PI178383 cold hardened at 2 C for 21 days and 49 days

(equal amount of cDNA pooled together before subtraction, tester) and subtracted against genotype Norstar cold hardened at 2 C for 1 day (24 H)(driver). Modified Smart cDNA (Clontech)priming and non-directional cloning"

ORIGIN	Query Match	3.3%;	Score 80.6;	DB 14;	Length 869;
	Best Local Similarity	44.9%;	Pred. No. 2.7;	Indels 0;	Gaps 0;
	Matches	305;	Conservative 0;	Mismatches 374;	Indels 0;
QY	580	AAGAAGCCATCGGACCGGTGTGATGCGGCCCAACCCCGTGAACATATCGCGCGCAAC	639		
DB	827	ATGGCGCGCGCGCAACACGAGGATCATCAACAGGAGGAGAGCAACCAACCAACCAAC	768		
QY	640	ATGCTGACCCAGTGGGCTGACCCCTGAACTTCCCATCGACCCCATCGAGCCGTCGCC	699		
DB	767	CACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC	708		
QY	700	GTGAAGCTGAAGCCCGGATGAGCGGCCCAACAGGTGAAGCAGTGGCCCTGACCGGAG	759		
DB	707	ACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC	648		
QY	760	AAGTATAGGCGCTGACCGCTCTCGGAGGATGGAGGAGGAGGAGGAGGAGGAGGAG	819		
DB	647	AACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC	588		
QY	820	ATCGGCCCGGAGAACCCCTACACACCCCGCTGTTCCCATCAAGAGAGGAGGAGGAG	879		
DB	587	AACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC	528		
QY	880	AAGTGGCGGAGGTGTGAGTTCGCGGAGCTGAACAGGCGACCCAGGACTTCTGGGAG	939		
DB	527	AACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC	468		
QY	940	GTGAGTGGGCTTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	999		
DB	467	AACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC	408		
QY	1000	GAGTGGGCGAGCGCTTCTTACGCTGTCCTCGGAGGAGGAGGAGGAGGAGGAGGAG	1059		
DB	407	GACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC	348		
QY	1060	TTCAACATCCCGAGCTACACACGAGACCCCGGAGTCCGCTACCAAGGAGGAGGAG	1119		
DB	347	GACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC	288		
QY	1120	CCCGAGGCTGGAAGGCGAGCCCGAGATTTCCAGAGGAGGAGGAGGAGGAGGAGGAG	1179		
DB	287	AACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC	228		
QY	1180	CCCTTCGCGCGCGACCCCGAGTGTGATCTACCGAGGCGCCCGCTGTACGTTGGCAG	1239		
DB	227	AACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC	168		
QY	1240	GACCTGGAGATCGGCCAGC	1258		
DB	167	AACGGCAACCAACCAAC	149		

RESULT 4  
BM320864  
LOCUS  
DEFINITION  
Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA  
1132 bp mRNA linear EST 03-JAN-2002  
rockefeller.0.46 Mastigamoeba balamuthi lambda ZAP II Library  
Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA  
sequence.  
BM320864  
ACCESSION  
VERSION  
BM320864.1 GI:18055270  
KEYWORDS  
EST.  
SOURCE  
Mastigamoeba balamuthi  
ORGANISM  
Mastigamoeba balamuthi  
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.  
REFERENCE  
1 (bases 1 to 1132)  
AUTHORS  
Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Senses,C.W.,

Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and Philippe,H.  
The analysis of 100 genes supports the grouping of three highly divergent amoebeae: Dictyostelium, Entamoeba, and Mastigamoeba  
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
21819461  
11830664  
Contact: Muller Miklos  
Laboratory of Biochemical Parasitology  
The Rockefeller University  
1230 York Avenue, New York, NY 10021, USA  
Email: mmuller@rockefeller.edu  
Insert Length: 1132 Std Error: 0.00  
POLYA=No.  
Location/Qualifiers  
1. .1132  
/organism="Mastigamoeba balamuthi"  
/mol\_type="mRNA"  
/strain="ATCC 30984"  
/db\_xref="taxon:108607"  
/clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
/note="syn: Phreatamoeba balamuthi"

ORIGIN	Query Match	3.3%;	Score 80.2;	DB 12;	Length 1132;
	Best Local Similarity	45.3%;	Pred. No. 3.1;	Indels 9;	Gaps 2;
	Matches	375;	Conservative 0;	Mismatches 444;	Indels 9;
QY	139	CGCGCCCGCGCAAGAGGCGTCTGGAAGTGGCGGCAAGAGGCGCCACAGATGAGGAC	138		
DB	79	CGCGCGAGGCGCAAGACGAGTACCGCGCGGCCNCTGGTGATCCAGGCAAGAACAG	138		
QY	199	TGCAACGAGGCGCAAGGCAACTTCTTCCGCGAGAGACTGGCCCTTCCCGAGGCGAGGCC	258		
DB	139	TACAACAGCCCCAGTAGTACCGCTTCGTC-----GTCCGCTTCCCAACAGGAGCATGTC	132		
QY	259	CGCGAGTTCGCCAGCGAGCAGACCGCGCCCAACAGCGCCACAGCGCGCGAGTGCAGGTG	318		
DB	193	TGCCAGATCGCTACGCGCAAGATCGAGCGGACCAATCTCTCGCGCGCGCTACTCGCAC	252		
QY	319	CGCGGCGACAAACCCCGCGAGCGCGCGCGAGCGCGCGAGGCGCGAGGCGAGCTTCCCT	378		
DB	253	GAGCTCACCGCTTTCGCGCGTCAAGCTCGGCTGACCAACTACCGCGCGCTACGCGACT	312		
QY	379	CAGATCACCTGTGGAGCGCCCGCTGGTGAGCATCAAGGTGGCGCGCGAGATCAAGGAG	438		
DB	313	GGCTGTGCTGGCGCGCGCTGGTGCTGAAGAGCTCACTCGACTCCAAAGTAGGAGGT	372		
QY	439	GCCCTGTGAGACACCGCGCGCGAGCACCCGTGTGTGAGAGAGATGAGCTTCCCGGCAAG	498		
DB	373	GTCAAGAAGGTCAACGCGCGAGGACTACAAGTCTGAGGAGCTCGACGAGCGCGCGCGT	432		
QY	499	TGGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCCAGTACGACAG	558		
DB	433	TTCAGGCGCTGTCTGAGCTCGGCTGGTCCGACCTTCGACTGGCGCGCGCTGTCGCC	492		
QY	559	ATCCTGATCGAGATCTCGGCGCAAGAGGCCATCGGACCGCTGTGATCGGCCCAACCCCC	618		
DB	493	GCCTCAAGGCGCATGTGCGACGCGCGGTCAACGTCCCC---CACAGCGAGACCGCGTTC	549		
QY	619	GTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAGACTTCCCATC	678		
DB	550	GTCCGCTTCAACGCGCGCAAGAGAGGAGCTCAACCGCGCGTCTTCCCAAGTACATCTTC	609		
QY	679	AGCCCATCGAGACCGTGGCGCTGAAGCTGAAGCGCGCGCATGGAGCGGCCCAAGGTGAAG	738		
DB	610	GGCGGCGACGTCGCGCGGTACATGAAGCTCTCAAGGAGCGAGGAGCGCGCGCTTCGAC	669		
QY	739	CAGTGGCGGCTGACCGGAGGAGAGATCAAGGCGCTGACCGCATCTCGAGGAGATGGAG	798		
DB	670	CGCAGATTCTCGCGCTACGCGCAAGAGGAGGTGTCTACCGCGGACATGCTCGAGAGATAC	729		
QY	799	AAGGAGGCGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTACACACCCCGCTGTCGC	858		

```

730 ACCGAGGCCCAAGCAGATCCGCGCGACCCGACCTTCGTCCCAAGCGCGCTCGAAG 789
859 ATCAAGAAGAGGACAGACCAAGTGGCGGAAGTGGTGGAGTTCCTCCGCGAGCTGAACAAG 918
790 CCGAGGGCCCAAGCCAGCATCGGGCAGCGAGGCTGAGTACACGAGCGCAAG 849
919 CGCACCCAGGACTTCTGGAGGTCAGCTGGCCATCCCCCAACCCCGCC 966
850 AACCGCGTCGCCAGAGAAGGTCGCGCTGGCTACCCCGACGCCCC 897

RESULT 5
BM320900 1165 bp mRNA linear EST 03-JAN-2002
LOCUS
DEFINITION
rockefeller.0.353 Mastigamoeba balamuthi lambda ZAP II Library
Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
sequence.
ACCESSION
BM320900.1 GI:18055306
VERSION
Mastigamoeba balamuthi
KEYWORDS
Mastigamoeba balamuthi
SOURCE
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
ORGANISM
1 (bases 1 to 1165)
REFERENCE
Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Seneen, C.W.,
Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and
Philippe, H.
The analysis of 100 genes supports the grouping of three highly
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
21819461
11830664
PUBMED
COMMENT
Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 1165 Std Error: 0.00
POLYA=No.

FEATURES
source
1..1165
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/note="syn: Phreatamoeba balamuthi"

ORIGIN
Query Match 3.2%; Score 79.6; DB 12; Length 1165;
Best Local Similarity 45.1%; Pred. No. 3.7;
Matches 428; Conservative 0; Mismatches 506; Indels 16; Gaps 3;

90 CATCAAGTCTTCAACTGCGGCAAGGAGGCGCACATCGCCGCAACTCGCGGCCCCCG 149
46 CGTCAAGAACAAAGGGGTACTTCAAGCGTCTCCAGACCCAGTTCCTCGCCCGCGAGGG 105
150 CAAGAGGCGCTCTGAGTGGCGGCAAGAGGCGCCACAGATGAAGGACTGCACCGAGCG 209
106 CAAGACGCACTACCGCGCGCCANCTGTGTATCCAGACAGAACAAAGTACACAGCCC 165
210 CCAGGCCAATCTTTCCTCGGAGGACTGCGCTTCCTCCAGGCGCAAGGCGCGAGTTCC 269
166 CAAGTACCGCTTCGTC-----GTCCGCTTCACCAACAGGACATCGTCTGCAGATCGC 219
270 CAGCGAGCAGACCGCGCCCAAGCCACAGCCCGCGAGCTGCAAGTGGCGGCGACAA 329
220 CTACGCCAAGATGACGGCGACACATCTCTCGCGCGCTTACTCGCAGAGCTCACCG 279
330 CCCCCCAGCGAGCGCGCGCGAGCGCCAGGCGACCTTGAACTTCCCTCCAGATCACCT 389
280 CTTGCGCGTCAAGCTCGCTGCTGACCAACTACGCGCGCGCTACGCGACTGCGCTGCTGCT 339

```

```

390 FTGGCAGCGCCCCCTGTGTGAGCATCAAGTGGCGCGCCAGATCAAGGAGGCGCTGCTGGA 449
340 GCGCGCGCTGTGTGTGAAGAAGCTCAACCTCTACTTCCAAAGTACGAGGGTGTCAAGAAGGT 399
450 CACCGCGCGCGACACACCGTGTCTGGAGGAGATGAGCTGCGCGCAAGTGAAGAGCCCAA 509
400 CAACGCGGAGGACTAAGCTGAGAGAGCTGACGACGGCGCGCGCTTCAAGGCC-- 457
510 GATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCGCAGTACGACCGAGATCTCTATCGA 569
458 --TGCTCGACGTGCGCTGCTCGCACCTCGACTGCGCGCGCGCTGTTTCGCGCGCTCAA 515
570 GATCTCGCGCAGAAGGCCATCGGCACCGTGTGATCGCGCGCCACCCCGCTGAACATCAT 629
516 GGGCATGTGCGACGCGCGGTCAAGTCCCCACAGCAGACCCGCTTGTGCGGTTCAA 575
630 CCGCGCGCAACATGCTGACCCAGCTGCGGTGCAACCTGAACTTCCCATCAGCCCATCGA 689
576 CCGCGCACAGAAGGAGCTCAACGCGCGGTCTCCGCAAGTACATCTTCGCGCGCAAGT 635
690 GACCGTGGCGGTGAGTGAAGCCCGGATGACGCGCGCCCAAGGTGAAGCAGTGGCCCT 749
636 CCGCGGTATCATGAAGCTCTCAAGGAGCAGACGCGCGCTTCGACCGCGCAAGTCTTC 695
750 GACCGGAGGAGAAGATCAAGGCGCTGACCGCGCATCTGCGAGGAGATGGAAGAGGCGCAA 809
696 GCGCTACGCCAAGGAGGTGTCAACGCGGACATGCTCGAGAAGATCTACACCGAGGCCA 755
810 GATCACAAGATCGGCGCGCGAGAACCCCTAACAACCCCGCTTTCGCCATCAAGAGAA 869
756 -----CAAGCAGATCCGCGCGCAGCCGACCTTCTGTCGCCAAGCGCGCTCGAAGCCGA 809
870 GGACAGCACCAAGTGGCGCAAGCTGTGTGACTTCGCGAGCTGAAACAAGCGCACCCAGGA 929
810 GGGCGCCAGCCCAAGCACTTGGGGCAGCGAGGTGAGTACCAGCGCGCAAGAACCG 869
930 CTTCTGGGAGGTGACGTGGGATCCCCCACCACCGCGCGCTTGAAGAGAAAGAGCGT 989
870 CGTGGCGCAGAGAAGGTCCGCTGCGCTACCGCGCGCGCGCTCGAAGCGCGCA 929
990 GACCGTGTGGAGCTGGCGACGCTTCTGAGCGTGGCGCGCTGGAGCGG 1039
930 CGTGGCGCTGTACACCGCATCTTCGCGTGGCGGTTCGCTGCTGCGG 979

```

```

RESULT 6
BM321430 867 bp mRNA linear EST 03-JAN-2002
LOCUS
DEFINITION
rockefeller.0.1153 Mastigamoeba balamuthi lambda ZAP II Library
Mastigamoeba balamuthi cDNA similar to ribosomal protein S4, mRNA
sequence.
ACCESSION
BM321430
BM321430.1 GI:18055836
VERSION
EST.
KEYWORDS
Mastigamoeba balamuthi
SOURCE
Mastigamoeba balamuthi
ORGANISM
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE
1 (bases 1 to 867)
AUTHORS
Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Seneen, C.W.,
Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and
Philippe, H.
The analysis of 100 genes supports the grouping of three highly
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
21819461
11830664
PUBMED
COMMENT
Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 867 Std Error: 0.00

```

```

FEATURES             Location/Qualifiers
     source           1..867
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"

ORIGIN
Query Match       3.2%; Score 79; DB 12; Length 867;
Best Local Similarity 45.7%; Pred. No. 4.1;
Matches 314; Conservative 0; Mismatches 370; Indels 3; Gaps 1;

QY 58 CAGGCGCAACTTCAAGGGCCCAAGCGGATCATCAAGTCTTCAATCGGCAAGGAG 117
DB 131 CGGACAAAGATGCGAGTGCCTCGGTCTATCTCTGTCGCAACAGGTTGAAGTAC 190

QY 118 GGCACATCGCCGCACTGCGGCCCGCCCGCAAGAGGGTCTGGAAGTGGCGAAG 177
DB 191 GCGGTGACCCCGCGTGAAGTCACTCGATCGTGTGATGAGCGGCTGATCAAGATCGAGGC 250

QY 178 GAGGCGCACCAAGATGACCGGATGACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTG 237
DB 251 AAGTCCGACCGACAGCACTTCCCGCGGCTTCATGACGTCGTCGATCGACAAAG 310

QY 238 GCCTTCCCGCAGGCGCGGCGAGTTCGCCAGCGAGCGAAACCGCGCCCAAGAGCCC 297
DB 311 ACCGACGAGCACTTCCGCTCTCTAGACACCAAGAGGCGCTTCCAGGGCGACCGCATC 370

QY 298 ACCAGCCGCGAGTSCAGGT--GGCGCGGACAAACCCCGCAGCGAGGCGCGCGCGAG 354
DB 371 AACTCGACCGAGGCCAAGTTCAAGCTCGCAAGTCCCGCGCGTGCAGCTCGGCAACAAG 430

QY 355 CGCCAGGCGACCTTGAACTTCCCGCAGATACCTGTGGAGCGCGCCCTGTGGTAGCATC 414
DB 431 GGCATCCCGTACTCTGTGACCCACAGCGCGGCGACGATCCGCTACCCCAACCCCGCATC 490

QY 415 AAGTGGGCGGCGACATCAAGAGGCGCTGTGGACACCGCGCGGAGCACCGTGTG 474
DB 491 AAGTCAACGACACGCTCAAGATGACCTGTGCTCGGCGAGATCATCGATTCTGTGAG 550

QY 475 GAGGAGATGAGCTTCCCGCGCAAGTGGAGGCCCAAGATGATCGGCGGATCGGCGGCTTC 534
DB 551 TTCGAGATCGGCAACTCTGTATGATCACTGGCGGACGCAACCTTGGCGCGTCTGGGCTC 610

QY 535 ATCAAGTGCAGTACGACGATCTCTGATCGAGATCTGCGCAAGAGGCCATCGGC 594
DB 611 ATTGTGCGCGGAGAGCAGAGGCTCTGATCGAGATCATCGATCGGCAAGGCCACCAAGTCC 670

QY 595 ACCGTGCTGATCGGCGCCACCCCGTGAACATCATCGGCGCGCAACATCTGACCCAGCTG 654
DB 671 GGCACCAAGTTTCGCGAGCGGCTGACCAAGCTTCTGTGATCGGCAAGGCCACCAAGTCC 730

QY 655 GCGTGCACCTGAACTTCCCGATAGCCCGCATCGAGACGCGTCCCGTGAAGCTGAAGCCC 714
DB 731 CTCGTGACGCTCCCGCGCGGCAAGGCGCATCAAGAAAGTCAAGTCAATCGAGGAGTTCAGGCG 790

QY 715 GGCATGGAGCGGCCCAAGGTGAAGCAG 741
DB 791 CGCCAGCGGCACAGGACCGAGGAGGAG 817

RESULT 7
BM321022 1550 bp mRNA linear EST 03-JAN-2002
LOCUS rockefeller.0.1192 Mastigamoeba balamuthi lambda ZAP II Library
DEFINITION Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase [EC
3.3.1.1], mRNA sequence.
ACCESSION BM321022
VERSION BM321022.1 GI:18055428
KEYWORDS EST.

```

```

SOURCE             Location/Qualifiers
ORGANISM           1..1550
Mastigamoeba balamuthi
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE          1 (bases 1 to 1550)
AUTHORS            Bapteste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W.,
                  Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and
                  Philippe, H.
TITLE              The analysis of 100 genes supports the grouping of three highly
                  divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
                  Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
JOURNAL            21819461
MEDLINE            11830664
PUBMED             11830664
COMMENT            Contact: Muller Miklos
                  Laboratory of Biochemical Parasitology
                  The Rockefeller University
                  1230 York Avenue, New York, NY 10021, USA
                  Email: mmuller@rockefeller.edu
                  Insert Length: 1550 Std Error: 0.00
                  POLYA=No.

FEATURES            Location/Qualifiers
     source           1..1550
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"

ORIGIN
Query Match       3.1%; Score 76.6; DB 12; Length 1550;
Best Local Similarity 45.6%; Pred. No. 9;
Matches 308; Conservative 0; Mismatches 364; Indels 3; Gaps 1;

QY 1749 CGACCTGTAAGCAGCTGACCGAGGCGCTGCAGAAATGCCATGGAGACATCGTGATCTG 1808
DB 5 CGGCACGAGGACGAGGACCAACCGCGCGCGCCATCGCGAGGCGCGCTCGTCTGCTT 64

QY 1809 GGGCAAGACCCCAAGTTCGCTCCCTCCCATCCAGAGAGACCTGGGAGACTGGTGAC 1868
DB 65 CGCTGGAAGGCGAGAACTCCAGAGTACTGGAGTGCACCTGGAAGGCGCTGTGCTT 124

QY 1869 CGACTACTGGCAGGCGCACCCTGGATCCCGAGTGGGAGTTCGTGAACACCCCCCTGCT 1928
DB 125 CGGCCCTTACGAGGCGCTCAGATCATCGTCAAGCGGCGGTGACGCGACTCTCATGAT 184

QY 1929 GAAGCTGTGGTACCAAGCTGGAAGAGGCCATCATCGGCGCGGAGACCTTCTAGTGA 1988
DB 185 CCACAAGGGGTTCGCGCGCGAGGACAAACCCCAAGCTGTGGAGGACGAGGGGCTCGA 244

QY 1989 CGGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGGGCG 2048
DB 245 GGAAGTGGCTGCTCTCAACAGTGTCAAGAGTCCAGAAGGAGCAGCCCGGCTCTG 304

QY 2049 GCAGAAGATCGTGAAGCTTACCGAGACCAACCAAGAGACCGAGCTGCGAGGCCATCCA 2108
DB 305 GCACAAGATCTCTCCCGAGATCCGCGGTGTCAAGAGGAGAGACGAGTGGCGTGTATGAG 364

QY 2109 GCTGGCGCTTGCAGGACAGCGGAGGAGTGAACATCGTGACCGCAGCAGCTAGCCCT 2168
DB 365 GCTGTACCAAGTGCACCGGACCGCAAGCTGTGTTCCCGCGCTCAAGCTCAACGACTC 424

QY 2169 GGGCATCATCTCAGGCGCCAGCCGACAAAGAGCGAGAGCGAGCTGGTGAACAGATCATCA 2228
DB 425 TG---TCACCAAGAGCAAGTTTGACAAATCTACGGCTCGCGCACTCTCATCGACGG 481

QY 2229 GCAGCTGATCAAGAGAGAGAGTGTACTGAGTGGTGGCGCCCAAGGCGATCGG 2288
DB 482 CATCAAGCGCGGACCGACGATGATCTCGGCGCAAGGTCGCGCTGTCGCGGCTACGG 541

QY 2289 CGGCAACGAGCAGATGCACAAGCTGTGAGCAAGGGCATCCGCAAGGTGCTGTCTCTGA 2348
DB 542 CGAGCTGGGCAAGGGCTGCGCGCGAGTCTGCGCGCGCCAGGGGCTGCGCGCTCATCTGAC 601

```

QY	2349	CGGCATCGATGGCGGCATCGTGATCTTACACAGTACATGACGACCTGTACTGCTGGCAGCGG	2408
Db	602	GGAGATCGACCCCATCTGCGCGCTGCAGGCGTTCGATGGCGGGTTCGAGGTCAACACGCT	661
QY	2409	CGGCCCTAGGATCGA	2423
Db	662	CGAGGCGGGGCTCGA	676

RESULT 8	
BI724851	LOCUS
BI724851	linear EST 19-SEP-2001
DEFINITION	545 bp mRNA
	lamba3075805.y1 C. reinhardtii CC-1690, Stress ii (normalized),
	Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCSSION	BI724851
VERSION	BI724851.1
KEYWORDS	GI:15700546
SOURCE	EST.
ORGANISM	Chlamydomonas reinhardtii
	Chlamydomonas reinhardtii
	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
	Chlamydomonadaceae; Chlamydomonas.
REFERENCE	1. (bases 1 to 545)
AUTHORS	Groisman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C.,
	Lefebvre, P., Mobermott, J.P., Shrager, J., Sillow, C. and Stern, D.
TITLE	Analyses of the Chlamydomonas reinhardtii Genome: A Model,
	Universal System for Analyzing Gene Function and Regulation in
	Vascular Plants. Project: 1031
JOURNAL	Unpublished (2001)
COMMENT	Contact: Charles Hauser

```

FEATURES
source
1..545
    /organism="Chlamydomonas reinhardtii"
    /mol_type="mRNA"
    /strain="CC-1690 wild type mt+ 21gr"
    /db_xref="taxon:3055"
    /clone_lib="C. reinhardtii CC-1690, Stress II
    (normalized), Lambda Zap II"
    /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
    XhoI; Stress condition II library, constructed by John_
    Davies and Jeffrey McDermott, combines cDNAs from CC-1690
    cells grown to mid-log phase in TAP (NH4+ - containing)
    and shifted to TAP - NO3- (24hrs); H2 production
    conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant
    Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
    sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
    PolyA mRNA was purified from each sample, pooled and cDNA
    synthesized. The cDNA was directionally cloned into lambda
    Zap II (Stratagene) in the EcoRI (5') and XhoRI (3')
    sites. pBluescript II SK- plasmids were excised from the
    lambda Zap clones by superinfection with ExAssist
    (Stratagene) phage. The library was normalized using
    method 4 described in Bonaldo et al., (1996) Genome
    Research 6: 791-806."

```

ORIGIN									
Research 6: 791-806.									
	Query Match	3.1%	Score 75;	DB 12;	Length 545;				
	Best Local Similarity	46.7%;	Pred. No. 11;						
	Matches 237;	Conservative 0;	Mismatches 270;	Indels 0;	Gaps 0;				
Qy	1184	TCGCGGCCGCAACCCCGAGATCGTGATCTACACGAGCCCCCTCTACGTGGGCGACGACC	1243						
Db	36	TGCGCGCGACGACGCTGTTTGCCATCTCTGGAGGCCAGGTGGAGACGGCAACCCCT	95						
Qy	1244	TGGAGATCGGCGACGACCGCGCCCAAGATTCGAGGAGCTGCGCAACGACCTGCTGGCTGGG	1303						
Db	96	ACATGCTGTACAAGGACCACTTGCACCGCGAAGTCCAAACCGACGAACTCTGGGCACTCA	155						

1304	QY	GTCTACACACCCCGGACAGAAAGCACCAGAGAGAGCCCTCTCTGCCCACATCGAGCTGC	1363
1304	QY	GTCTACACACCCCGGACAGAAAGCACCAGAGAGAGCCCTCTCTGCCCACATCGAGCTGC	1363
156	Db	AGTGTCTCCAACTGTGTGACGAGATTATCGAGTACACACGACCCCGATGATCGCCGTGT	215
156	Db	AGTGTCTCCAACTGTGTGACGAGATTATCGAGTACACACGACCCCGATGATCGCCGTGT	215
1364	QY	ACCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGACACGTGA	1423
1364	QY	ACCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGACACGTGA	1423
216	Db	GCAACTGGCTCCATCGGCGCTTGCCTCCCGCTTGTGTGGAGCGCGGACGCCCGGCT	275
216	Db	GCAACTGGCTCCATCGGCGCTTGCCTCCCGCTTGTGTGGAGCGCGGACGCCCGGCT	275
1424	QY	ACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCA	1483
1424	QY	ACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCA	1483
276	Db	GGGAGAGCAAGAAGCTGGTGGGTGCTGTGAGCGCGCCAAACCGCTTCTTCGACTTCGAGA	335
276	Db	GGGAGAGCAAGAAGCTGGTGGGTGCTGTGAGCGCGCCAAACCGCTTCTTCGACTTCGAGA	335
1484	QY	AGGTGCGCAGCTGTGTGCAAGCTCTCTCGGCGGCCCAAGGCCCTGACCGGACATGTTGCCCC	1543
1484	QY	AGGTGCGCAGCTGTGTGCAAGCTCTCTCGGCGGCCCAAGGCCCTGACCGGACATGTTGCCCC	1543
336	Db	AGTGTGCTGAGATGACCAAGATGGTACCCGCAACCTGAAACAAGATCATTTGACGTCAACT	395
336	Db	AGTGTGCTGAGATGACCAAGATGGTACCCGCAACCTGAAACAAGATCATTTGACGTCAACT	395
1544	QY	TGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGCGGAGGCCGTGC	1603
1544	QY	TGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGCGGAGGCCGTGC	1603
396	Db	ACTATCCGCAACCCCAAGGCGCAGCGTCCAAATGCGCCACCCGCCCATCGGCTTGGCG	455
396	Db	ACTATCCGCAACCCCAAGGCGCAGCGTCCAAATGCGCCACCCGCCCATCGGCTTGGCG	455
1604	QY	ACGGCGTGTACTACGACCCCGACGAGGACCTGTGTGGCCGAGATCCAGAAGCAGGGCCACG	1663
1604	QY	ACGGCGTGTACTACGACCCCGACGAGGACCTGTGTGGCCGAGATCCAGAAGCAGGGCCACG	1663
456	Db	TGCAGGCGCTGGCGGACACTTTCATCTGTGTGGCATGGCTTCGACTCGCCGAGGCCG	515
456	Db	TGCAGGCGCTGGCGGACACTTTCATCTGTGTGGCATGGCTTCGACTCGCCGAGGCCG	515
1664	QY	ACCAGTGGACCTACCAAGATCTACAGG	1690
1664	QY	ACCAGTGGACCTACCAAGATCTACAGG	1690
516	Db	CCGAGCTGAACCCGACAGATCTTCGAG	542
516	Db	CCGAGCTGAACCCGACAGATCTTCGAG	542

RESULT 9  
EM321023  
LOCUS  
862 bp mRNA linear EST 03-JAN-2002  
rockefeller.0.594 Mastigamoeba balamuthi lambda ZAP II Library  
Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC  
3.3.1.1). mRNA sequence.

J. Mol. Evol. 60:1-17, March sequence.

ACCESSION  
EM321023

VERSION  
EM321023.1

KEYWORDS  
GI:18055429  
EST.

SOURCE  
Mastigamoeba balamuthi  
Mastigamoeba balamuthi  
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.

REFERENCE  
(bases 1 to 862)

AUTHORS  
Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,  
Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and  
Philippe,H.

TITLE  
The analysis of 100 genes supports the grouping of three highly  
divergent ancebae: Dictyostelium, Entamoeba, and Mastigamoeba

JOURNAL  
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)

MEDLINE  
21819461

PUBMED  
11830664

COMMENT  
Contact: Muller Miklos

COMMENT	FEATURES	source
<p>Laboratory of Biochemical Parasitology  The Rockefeller University  1230 York Avenue, New York, NY 10021, USA  Email: mmuller@rockvax.rockefeller.edu  Insert Length: 862 Std Error: 0.00  POLYA=No.</p>	Location/Qualifiers	1. .862

```

1.002
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/notes_syn: Phreatamoeba balamuthi"

```

ORIGIN	/notes=-Syn: FhritatanCeba paramutuu									
	Query Match	3.1%	Score 75;	DB 12;	Length 862;					
	Best Local Similarity	46.2%	Pred. No. 12;							
	Matches 283;	Conservative 0;	Mismatches 327;	Indels 3;	Gaps 1781					
Qy	1722	CGCCAAATCGGCACCGCCACACACGAGTGTAACGAGTGACCGGCCGTGCAGAA	1781							





14671302  
2 (bases 1 to 2433)  
PUBMED  
REFERENCE  
AUTHORS  
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarival, A.,  
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,  
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,  
Adams, M.D. and Cargill, M.  
Direct Submission  
TITLE  
JOURNAL  
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT  
This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
source  
Location/Qualifiers  
1..2433  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
<1..>2433  
/locus\_tag="HCN0810"  
gene  
ORIGIN  
Query Match 3.0%; Score 73.8; DB 29; Length 2433;  
Best Local Similarity 43.0%; Pred. No. 21;  
Matches 485; Conservative 0; Mismatches 632; Indels 12; Gaps 2;  
QY 463 GACACCGTCTGAGGAGATGAGCTGCCGCGAAGTGAAGCCCAAGATATCGCGGC 522  
Db GACAAGCTCCGGCAGCGCTCACCAGAACCGCGAGGCAAGCTGGAGCTGCACTGTTC 1299  
QY 523 ATCGCGCGGCTTCATCAAGGTGCGCCAGTACGACAGATCCTGATCGAGATCTGGGCAAG 582  
Db ATGCTCAGTGGCATCCCTGACACTGTGTTGACCTGTGGAGCTGGAGTCTCAAGCTG 1359  
QY 583 AAGGCCATCGGACCGTGTGATGCGGCCACCCCGTGAACATCATCGGCGGCAACATG 642  
Db GAGCTGTATCCCGAGCTGACCATCCGCGCCAGCATTTGCCAGCTCAGCGGCTCAAGGAG 1419  
QY 643 CTG-----ACCCAGCTGGGTGACCTGAACTTCCCATCAG-----CCCCATCGAG 690  
Db CTGTGGCTTACCAACAGCGGCCAAGATTGAAGCGCCCGGCTTCTTCTCGCGAG 1479  
QY 691 ACCGTGCGGTGAAGCTGAAGCCCGGATGAGCGGCCCAAGGTGAACATGTGCCCTG 750  
Db AACCTGCGGGGCTGCATCAATCAAGTTTACCAGCATCAAGGAGATCCCGCTGTGATCTAT 1539  
QY 751 ACCGAGGAGAGATCAAGCCCTGACCGCATCTCGAGGAGATGAAGAGAGGGCAAG 810  
Db AGCCTGAGACACTGGAGAGCTGCACCTGACGGCAACCTGAGCGGAGAAACAACCGC 1599  
QY 811 ATCAACAGATGCGGCCGAGAACCCCTTACAAACACCCCGCTGTTCCGCTCAAGAGAAG 870  
Db TACATCGTCACTGACGGGCTGCGGAGCTCAACCGCTCAAGGTGCTGCGGCTCAAGAGC 1659  
QY 871 GACACGACCAAGTGGCGCAAGCTGTGACTTCGCGAGCTGAACAGCGCACCCAGGAC 930  
Db AACCTAAGCAAGCTGCCAGGTGTACAGATGTGGCGGTGCACCTGCAAGAGCTGTCC 1719  
QY 931 TTCTGGAGGTGACGTGGGATCCCCCAGCCCGGCTGGAAGAGAAAGAGCGTG 990  
Db ATCAACAATGAGGCAACCAAGCTCATCTCTCAACAGCTCAAGAGATGGCGAACTG 1779  
QY 991 ACCGTGCTGGAGTGGGCGAGCCCTTACTTACGCTGCCCCCTGGAAGAGATTCCCGCAAG 1050  
Db ACTGAGCTGGAGCTGATCCGCTGTGACTGTGGAGCGCATCCCCACTCCATCTTCAGCTC 1839  
QY 1051 TACACCGGCTTACCATCCCGAGCATCAACAGAGACCCCGGATCCGCTACCAATAC 1110  
Db CACAACTCTCAGGAGATTGACCTCAAGGACAAACACCTCAAGACCATCGAGGATCATC 1899  
QY 1111 AACGTGCTGCCCGAGGCTGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATCAACCAAG 1170  
Db AGCTTCAGACACTGACCGCTCAGCTGCTTAACTGCTGAGTGAACCAACATCGCCTAC 1959  
QY 1171 ATCTGGAGCGCTTCGCGCGCCGCAACCCCGAGATCGTGAATCTACAGGCGCCCTGTAC 1230

1960 ATCCCATCATGATCGCAACCTCAACCACTGGAGCGCTTACTCTGAACCGCAACAG 2019  
1231 GTGGGAGGAGACCTGGAGATCGGCCAGCACCGGCCCAAGATGAGAGGTGGCGAAGCAC 1290  
2020 ATCGAGAAGATCCCAACCCAGCTCTTCTACTGCGCAAGCTGGCTTACCTGAGACCTCAGC 2079  
1291 CTGCTGCTGGGGCTTCAACCCCGGCAAGAGCACCAAGAGAGAGAGAGAGAGAGAGAG 1350  
2080 CACAAACACTGACCTTCTCTCTGCGGATCGGCTCTCTGAGAACCTCAGAACCTA 2139  
1351 CCATCGAGTGCACCCCGCAAGTGCAGCCGCTGAGCCCATCGAGTGCAGAGAGAGAG 1410  
2140 GCCATCAGCGCCAAACCGGATCGAGAGCTCCCTCCGAGCTCTTCAGTGCAGAGAGCTG 2199  
1411 AGCTGAGACCTGAACGACATCGAGAGCTGTGGGCAAGCTGAACCTGAGCGGAGAGATC 1470  
2200 CGGCGCTGACCTGCGGCAACACGCTGCTGAGTCACTCCCTCCAGGCTGGCGGAGCTG 2259  
1471 TACCCCGGATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGGCGGCCCAAGGCGCTGACC 1530  
2260 ACCAACCTGACGAGATCGAGCTGCGGCGCAACCGGCTGAGTGCCTGTGGAGCTG 2319  
1531 GACATGCTGCTGCTGACCGAGAGGCGGAGCTGAGCTGCGCGAGAGAGAGAGAGAG 1579  
2320 GCGAGTGGCCACTGCTCAAGCGCAGCGCTTGGTGGTGGAGAGAGAGAGAGAGAGAG 2368  
RESULT 12  
BM321393  
LOCUS  
DEFINITION  
rockefeller.0.1222 Mastigamoeba balamuthi lambda ZAP II library  
Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA  
sequence.  
ACCESSION  
BM321393  
VERSION  
BM321393.1 GI:18055799  
KEYWORDS  
EST  
SOURCE  
Mastigamoeba balamuthi  
Mastigamoeba balamuthi  
ORGANISM  
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.  
REFERENCE  
1 (bases 1 to 853)  
AUTHORS  
Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W.,  
Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and  
Philippe, H.  
The analysis of 100 genes supports the grouping of three highly  
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba  
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
21819461  
MEDLINE  
PUBMED  
COMMENT  
Contact: Muller Miklos  
Laboratory of Biochemical Parasitology  
The Rockefeller University  
1230 York Avenue, New York, NY 10021, USA  
Email: mmuller@rockefeller.edu  
Insert Length: 853 Std Error: 0.00  
POLYA=No.  
FEATURES  
Location/Qualifiers  
1..853  
/organism="Mastigamoeba balamuthi"  
/mol\_type="mRNA"  
/strain="ATCC 30984"  
/db\_xref="taxon:108607"  
/clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
/note="syn: Phreatamoeba balamuthi"  
ORIGIN  
Query Match 3.0%; Score 73.2; DB 12; Length 853;  
Best Local Similarity 46.9%; Pred. No. 20;  
Matches 299; Conservative 0; Mismatches 329; Indels 9; Gaps 2;  
QY 388 CTGTGGCAGCGCCCGCTGTGTGAGATCAAGTGGCGGCGAGATCAAGAGAGAGAGAGAG 447  
Db CTGGCGCGCGTGTGTGTGAAGAGAGTCAACCTCGACTCCAGATGAGAGAGAGAGAGAG 184

QY 448 GACACCGCGCGGACGACACCGTGTGAGGAGATGAGCTGCGCGGCAAGTGAAGCC 507  
 Db 195 GTCAACGCGGAGGACTACACGTCGAGGAGCTCGACGACGGGCGCCCGGTTCAAGGCC 244  
 QY 508 AAGATGATCGCGGCGCATCGCGGCTTCTCATCAAGTGTGCGCAGTAGTACGACAGATCCTGATC 567  
 Db 245 CTGCTCGACGTGCGGCTGTGCGACCTCGACCGCGCGCGCGGTGTCGCGCGCTCAAG 304  
 QY 568 GAGATCTCGCGCAAGAGCCATCGGACCGTGTGATGTCGCGCCCGCCACCGCGTGAACATC 627  
 Db 305 GGCATGTGCGACGCGCGGCTCAAGCTCCCC---CACAGCGAGACCCCGCTTCGTGCGGCTTC 361  
 QY 628 ATCGCGCCCAACATGCTGACCCAGCTGGGCTGCAACCTGCACTTCCCATCAGCCCATC 687  
 Db 362 AACGGGACAGAGAGAGCTCAACGCGCGGTCTCCGCAAGTACATCTTCGCGGCGCAC 421  
 QY 688 GAGACCGTGCCTGTGAGCTGAAGCCCGCGCATGACCGCCATCTCGGAGGATGGAAGAGGGGC 747  
 Db 422 GTGCGCGGTACATGAAGCTCCTCAAGAGGAGGAGCGCGCGCTTCGACCGCGCATTC 481  
 QY 748 CTGACCGAGGAGAGATCAAGGCGCTGACCGCCATCTCGGAGGATGGAAGAGGGGC 807  
 Db 482 TCGGCTACGCGAAGAGGGTGTACCGCCGACATGCTCGAGAGATCTACCGAGGCC 541  
 QY 808 AAGATCAACAGATCGGCGCGCGAGAACCCCTTACACACCCCGCTGTTGCGCATCAAGAG 867  
 Db 542 CA-----CAAGCAGATCGGCGCGACCCGACCTTCGTCCCAAGCGCGCTCGAAGCCC 591  
 QY 868 AAGCAGACACCAAGTGGCGCAAGCTGGTGACCTTCGCGAGCTGAACAGCGCACCCAG 927  
 Db 596 GAGGCGCCAGCCAGCATGCGGCGAGCGGAGCTGAGCTACCGAGCGCGCAAGAC 655  
 QY 928 GACTTCTGGGAGGTGACCTGGGCATCCCCCAGCGCGCGCTTGAAGAGAGAGAGAGC 987  
 Db 656 CGCGTGGCGCGAGAGAGTCCGCTGGGCTACCGCGAGCGCGCCCAAGAGCGCAAGTAATT 715  
 QY 988 GTGACCGTGTGGAGTGGCGGAGCGCTACTTCAGCG 1024  
 Db 716 CCGGTCGCGCTGTACACNGCCATCTCCGCTGCGGCG 752

RESULT 13  
 BM587428/c  
 LOCUS  
 DEFINITION 17000587321202 A.Gam.ad.cdna.blood1 Anopheles gambiae cdna clone  
 19600449696110 5', mRNA sequence.

BM587428 566 bp mRNA linear EST 25-FEB-2002  
 BM587428 19600449696110 5', mRNA sequence.  
 BM587428.1 GI:18883289  
 EST.  
 Anopheles gambiae (African malaria mosquito)  
 Anopheles gambiae  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 Anopheles.

REFERENCE  
 AUTHORS Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L.,  
 Charlab, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.  
 Celeris Anopheles gambiae EST project  
 Unpublished (2002)  
 Contact: Holt R.A.  
 Celeris Genomics  
 45 W. Gude Dr., Rockville, MD 20850, USA  
 Tel: 2404533151  
 Fax: 2404534580  
 Email: HoltR@celera.com  
 Plate: NU01004AX row: G column: 08  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1. .566  
 /organism="Anopheles gambiae"  
 /mol\_type="mRNA"  
 /strain="RSP-ST (Reduced susc. to Permethrin - std.

FEATURES  
 source

chromosome)"  
 /db\_xref="taxon:7165"  
 /clone="19600449696110"  
 /dev\_stage="Adult"  
 /lab\_host="DH10b"  
 /clone\_lib="A.Gam.ad.cdna.blood1"  
 /vector="psort1; Site 1; Sal1; Site 2; Not1; Whole  
 adult mosquitoes (mixed sex) frozen on liquid nitrogen 24  
 hours after human blood feeding. cdna inserts >500 bp  
 cloned directionally into psort 1. Not 1 site is 3'.  
 Clones available through the Malaria Research and  
 Reference Reagent Resource Center (www.malaria.mr4.org)"

## ORIGIN

Query Match 3.0%; Score 72.8; DB 12; Length 566;  
 Best Local Similarity 48.3%; Pred. No. 21;  
 Matches 203; Conservative 0; Mismatches 217; Indels 0; Gaps 0;  
 QY 1944 GTTGAGAGAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGAGCGCGCGCAACCG 2003  
 Db 482 GTTCGAGATCGACGCAACGGCATCTCGAGTGTGCGCCGAGGACAAAGGCGACGGGCAA 423  
 QY 2004 CGAGACCAAGATCGCAAGGCGGTAGTGAACCGCGCGCGCGCGAGAGATCGTAG 2063  
 Db 422 CCGGAAAAGATCGTCATCAACGACGAGAACCGCTGACGCGGAGACATCGAGCG 363  
 QY 2064 CCGGACCAAGACCAACCAAGAGACGAGCTGCGAGCGCATCCAGCTGGCGCTGCGAGA 2123  
 Db 362 CATGATCAAGGATCGGAGCGGTTCGCCGACGACGACAAAGAGCTGAAGGAGCGGCTGA 303  
 QY 2124 CAGCGGAGCGAGGTGAACATGTCGACGACGACGACGACGACGACGACGACGACGACG 2183  
 Db 302 GCGCGCGACGACGAGCTCGAGAGCTACGCGTGAAGAACCAAGCTCAGCTCGAGTGAAGGA 243  
 QY 2184 CCAGCGCGACAGAGAGGAGGAGCTGTGAACAGATCATCGAGAGCTGATCAAGAA 2243  
 Db 242 CAAGCTGGCGGAGGAGTGTCCGACGACGACGACGACGACGACGACGACGACGACGACG 183  
 QY 2244 GGAGAGGTGTACCTGAGCTGGGTCCCGCCCAAGGCGATCGGCGCGACGAGCAGAT 2303  
 Db 182 GAAGATCAAGTGGCTGGACGACGACGACGACGACGACGACGACGACGACGACGACG 123  
 QY 2304 CGACAGCTGTGAGGAGGCGATCCGCAAGGTGTCTTCTGAGCGCATCGATGGCGG 2363  
 Db 122 GAAGAGCTGGAGGACATCTGTCAGCCCATCATCTCCAGCTGTACGCGAGCAGTGGCGG 63

RESULT 14  
 CB643171  
 LOCUS  
 DEFINITION

CB643171 788 bp mRNA linear EST 08-APR-2003  
 OSJNEB03L13.f OSJNEB Oryza sativa (japonica cultivar-group) cdna  
 clone OSJNEB03L13 5', mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

OSJNEB03L13.f OSJNEB Oryza sativa (japonica cultivar-group)  
 Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 1 (bases 1 to 788)  
 Jantasuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,  
 Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.  
 Large-scale identification of ESTs involved in the interaction  
 between rice and Magnaporthe grisea  
 Unpublished (2003)  
 Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288

Email: <http://genome.arizona.edu>

PCR Primers

FORWARD: gta aaa cga cgg cca gtg

BACKWARD: sga aac agc tat gac cat g

Plate: 03 row: L column: 13

Seq primer: gta aaa cga cgg cca gtg.

Location/Qualifiers

1..788

organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="rRNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/clone="OSJNE03113"

/tissue type="Leaf"

/dev stage="3 week"

/lab\_host="DH10B"

/clone\_lib="OSJNEB"

/note="Vector: plasmidscript II KS +; Site 1: EcoRI; Site 2:

XhoI; 24 hrs after inoculation with Rice Blast (Che

86061)"

# ORIGIN

Query Match 3.0%; Score 72.6; DB 14; Length 788;  
Best Local Similarity 46.3%; Pred. No. 24;  
Matches 314; Conservative 0; Mismatches 355; Indels 9; Gaps 2;

Qy 107 GCGCAAGAGGCGCAATCCCGCAACTCCCGCGCCCGCCGCAAGAGGGCTGTGGA 166

Db 99 GCGCGGCGCGCGCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 158

Qy 167 AGTGGCGAGAGGGGCCACAGATGAGGACTCCAGGAGCGCGCGCGCGCGCGCG 226

Db 159 AGTTCCGCGAGCGCGCTGGTCAAGATCCAGGCGCGCGCGCGCGCGCGCGCG 218

Qy 227 GCGAGGACCTGGCTTCCCGCAGGCGCAAGCGCGCGCGCGCGCGCGCGCGCG 286

Db 219 CCGTCGCCAGGCGCAAGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 278

Qy 287 CCAACAGCCCGACAGCGCGGAGCTGCGGTGGCGCGCGCGCGCGCGCGCGCG 346

Db 279 CCAGCGTCAAGGCGAGCGAGTGGATCTCAACTGCATCGCGCGCGCGCGCGCG 338

Qy 347 GCGCGGAGCGCGAGGCGACCTGAACTTCCCGCAGATCACTTGTGCGAGCGCGCG 406

Db 339 AGGCGTCAACCGGCTTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 398

Qy 407 TGAGCATCAAGTGGCGCGCGAGTCAAGAGGCGCGCGCGCGCGCGCGCGCGCG 463

Db 399 TCCAGTCAAGTGGCGCGCGAGTCAAGAGGCGCGCGCGCGCGCGCGCGCGCG 458

Qy 464 ACACCGTGGTGGAGGATGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 523

Db 459 ACAGCTGCGTGGAGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 518

Qy 524 TCGCGCGCTTCAAGGTGCGCGAGTACGACCGAGATCTCTGATCGAGATCTGCG 583

Db 519 AGGCTACTCGGCGATCCGGTTGAGATCTCGAGGCGCGCGCGCGCGCGCGCG 578

Qy 584 AGGCGATCGGCGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 643

Db 579 GCGTCAGCGCGTGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 638

Qy 644 TGACCGAGTGGGTGACCTGAACTTCCCGATCAGCGCGCGCGCGCGCGCGCG 703

Db 639 TGTCTACATTGCGGCGCTCATACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 698

Qy 704 AGCTGAAGCGCGGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 763

Db 699 GCAGGAG-----GTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 752

Qy 764 TCAAGGCGCGTGAACCGCA 781

Db 753 TCTTACGCTGAACCGCA 770

## RESULT 15

CC675888

LOCUS

DEFINITION

CC675888

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Ze mays

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CC675888 753 bp DNA linear GSS 19-JUN-2003  
OGWCO51TH ZM\_0.7 1.5 KB Zea mays genomic clone ZMWEMA0539J05,  
genomic survey sequence.

CC675888

CC675888.1 GI:32080584

GSS.

Ze mays

Ze mays

Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

ciade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 753)

White, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
Citek, R.W., Nunberg, A., Robbins, D. and Lake, N.

Consortium for Maize Genomics

Unpublished (2002)

Other\_GSSs: OGWCO51TV

Contact: Cathy White

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitecaw@tigr.org

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1..753

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMWEMA0539J05"

/clone\_lib="ZM 0.7 1.5 KB"

/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"

## FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches 281; Conservative

0; Mismatches 321; Indels 6; Gaps 1;

Qy 31 CAGGCGACGCGCGCAATCCTGATGCGAGGCGCGCGCGCGCGCGCGCGCGCG 90

Db 40 CACACCG 99

Qy 91 ATCAAGTCTTAACTGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 150

Db 100 GTCGACCG 159

Qy 151 AAGAAGGCTGCTGGAAGTGGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCG 210

Db 160 GACGAGCG 219

Qy 211 CAGGCGCACTTCTTCCGCGAGGAGCTGGCGCTTCCCCCAGGCGCGCGCGCGCG 270

Db 220 TACATCCGCGACCTCATCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 279

Qy 271 AGCGAGCAGAACCG 330

Db 280 CTCTACACCAACCG 339

Qy 331 CCCGCGAGCG 390

Db 340 GCCTTCG 399

Qy 391 TGGCAGCG 450

Db 400 GTCG 459

Search completed: April 10, 2004, 16:16:27  
Job time : 4130.92 secs







QY 2041 CGGGGCGGCGAGAGATCGTGTGACCTGACCGAGACCAACAGAGACCGAGCTGCAG 2100  
 Db CGGGGCGGCGAGAGATCGTGTGACCTGACCGAGACCAACAGAGACCGAGCTGCAG 2100  
 QY 2101 GCATCCAGCTGGCCCTGCGAGGACGCGGAGGAGTGAACATCGTGACCGAGCCAG 2160  
 Db GCATCCAGCTGGCCCTGCGAGGACGCGGAGGAGTGAACATCGTGAGCCAGCCAG 2160  
 QY 2161 TAGCCCTGGGCGATCATCCAGGCGCCAGCCGCAAGAGGAGAGCGAGTGTGAACCCAG 2220  
 Db TAGCCCTGGGCGATCATCCAGGCGCCAGCCGCAAGAGGAGAGCGAGTGTGAACCCAG 2220  
 QY 2221 ATCATCGAGCAGCTGATCAGAGAGGAGAGTGTGACTGAGCTGGTGGCCGCCCAAG 2280  
 Db ATCATCGAGCAGCTGATCAGAGAGGAGAGTGTGACTGAGCTGGTGGCCGCCCAAG 2280  
 QY 2281 GGATCGGGCGCAACGAGCAGATCGACAAGCTGGTGAAGGCGCATCCGAGGTGCTG 2340  
 Db GGATCGGGCGCAACGAGCAGATCGACAAGCTGGTGAAGGCGCATCCGAGGTGCTG 2340  
 QY 2341 TTCTTGACGGCATCGATGGGGCATCGTATCTACAGTACATGACGACCTGTACGTG 2400  
 Db TTCTTGACGGCATCGATGGGGCATCGTATCTACAGTACATGACGACCTGTACGTG 2400  
 QY 2401 GGACGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457  
 Db GGACGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457

RESULT 2  
 ACA03546  
 ID ACA03546 standard; DNA; 2445 BP.  
 AC ACA03546;  
 DT 22-MAY-2003 (first entry)  
 DE Synthetic DNA encoding immunogenic HIV peptide #29.  
 KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
 KW gene therapy; packaging cell line; humoral immune response;  
 KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
 OS Synthetic.  
 PN WO2003004657-A1.  
 PD 16-JAN-2003.  
 PF 05-JUL-2002; 2002WO-US021421.  
 PR 05-JUL-2001; 2001US-0303192P.  
 PR 31-AUG-2001; 2001US-0316860P.  
 PR 16-JAN-2002; 2002US-0349728P.  
 PR 16-JAN-2002; 2002US-0349793P.  
 PR 16-JAN-2002; 2002US-0349871P.  
 PA (CHIR ) CHIRON CORP.  
 PI Zur Megede J, Barnett SW, Lian Y;  
 DR WPI; 2003-221602/21.  
 XX New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
 PT polypeptides, useful as immunogenic compositions or vaccines for  
 PT generating humoral or cellular immune responses against HIV in a subject,  
 PT especially humans.  
 XX Example 1; Fig 34; 262pp; English.  
 PS The invention describes a synthetic polynucleotide encoding 2 or more  
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
 CC derived from different HIV subtypes. The polynucleotide is useful for

CC immunisation, generation of packaging cell lines, or production of HIV  
 CC polypeptides. The polynucleotide and its encoded proteins are useful as  
 CC immunogenic compositions or vaccines for generating humoral or cellular  
 CC immune responses against HIV in a subject, or for inducing neutralising  
 CC antibodies against HIV. The gene delivery vector comprising the  
 CC polynucleotide is also useful for DNA immunisation of, or for generating  
 CC an immune response (e.g. a humoral or cellular immune response) in, a  
 CC subject such as a mammal, particularly a human. This sequence encodes a  
 CC human immunodeficiency virus immunogenic peptide  
 XX  
 SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;  
 Query Match 99.4%; Score 2443.4; DB 7; Length 2445;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-294;  
 Matches 2444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 7 GCCACATGCGCGAGGCGCATGAGCCAGGCGCACCGAGGCGCAACATCTGTGACGCGGAGC 66  
 Db 1 GCCACATGCGCGAGGCGCATGAGCCAGGCGCACCGAGGCGCAACATCTGTGACGCGGAGC 60  
 QY 67 AACCTCAAGGGCGCCCAAGCGCATCATCAAGTGTTCCTCACTCGGCAAGGAGGCGCACATC 126  
 Db 61 AACCTCAAGGGCGCCCAAGCGCATCATCAAGTGTTCCTCACTCGGCAAGGAGGCGCACATC 120  
 QY 127 GCCCGCAACTGCGCGCGCCCGCGCAAGAGGGGTGTGGAAGTGTGGCAAGGAGGCGCAC 186  
 Db 121 GCCCGCAACTGCGCGCGCCCGCGCAAGAGGGGTGTGGAAGTGTGGCAAGGAGGCGCAC 180  
 QY 187 CAGATGAAGGAGTGCACCGAGGCGCGCAAGTGTTCCTCGCGAGGAGCTGCGCTTCCCC 246  
 Db 181 CAGATGAAGGAGTGCACCGAGGCGCGCAAGTGTTCCTCGCGAGGAGCTGCGCTTCCCC 240  
 QY 247 CAGGGCAAGGGCGCGAGTGTCCCGAGGAGCAGAACCGCGCCCAACAGCGCCACACGCGCG 306  
 Db 241 CAGGGCAAGGGCGCGAGTGTCCCGAGGAGCAGAACCGCGCCCAACAGCGCCACACGCGCG 300  
 QY 307 GAGCTGCAAGTGTGCGGCGGCAACCCCGAGCGAGGCGCGCGCGAGCGCGCGCGCGCGCG 366  
 Db 301 GAGCTGCAAGTGTGCGGCGGCAACCCCGAGCGAGGCGCGCGCGAGCGCGCGCGCGCGCG 360  
 QY 367 CTGAACTTCCCCCAGATCACCTGTGTGCGAGCGCGCGCGCTGTGTGAGCATCAAGGTGGCGCG 426  
 Db 361 CTGAACTTCCCCCAGATCACCTGTGTGCGAGCGCGCGCGCTGTGTGAGCATCAAGGTGGCGCG 420  
 QY 427 CAGATCAAGAGGCGCGCTGTGTGACACCGCGCGCGAGCAGACCGCTGTGGAGAGATGAGC 486  
 Db 421 CAGATCAAGAGGCGCGCTGTGTGACACCGCGCGCGAGCAGACCGCTGTGGAGAGATGAGC 480  
 QY 487 CTGCGCGGCAAGTGTGAAGCGCCCAAGATGATCGCGCGGCGATCGCGCGGCTTCATCAAGGTGCGC 546  
 Db 481 CTGCGCGGCAAGTGTGAAGCGCCCAAGATGATCGCGCGGCGATCGCGCGGCTTCATCAAGGTGCGC 540  
 QY 547 CAGTACGACAGATCTGTGATGAGATCTGTGGCAAGAGAGGCGCATCGGACCGTGTGTATC 606  
 Db 541 CAGTACGACAGATCTGTGATGAGATCTGTGGCAAGAGAGGCGCATCGGACCGTGTGTATC 600  
 QY 607 GCGCGCGCGCGCGTGAACATCATCTGCGCGCGCAACATGTGTGACCGAGCTGGGTGCAACCGCTG 666  
 Db 601 GCGCGCGCGCGCGTGAACATCATCTGCGCGCGCAACATGTGTGACCGAGCTGGGTGCAACCGCTG 660  
 QY 667 AACCTTCCCCCATCGAGCG 726  
 Db 661 AACCTTCCCCCATCGAGCG 720  
 QY 727 CCCAAGGTGAAGCAGTGGCG 786  
 Db 721 CCCAAGGTGAAGCAGTGGCG 780  
 QY 787 GAGGAGATGAGAGAGGCGCGAGATCCACAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 846  
 Db 781 GAGGAGATGAGAGAGGCGCGAGATCCACAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840  
 QY 847 CCGGTGTTCGCCATCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 906

Db 841 ||||| CCCTGTTTCCCATCAAGAGAGGAGCAGCACCAGTGGCGAAGTGGTGGACTTCGCG 900  
 Qy 907 GAGCTGAAACAAGCGCAACCCAGGACTTCTGGAGAGTTCAGTGGGATCCCCACCCCGCC 966  
 Db 901 GAGCTGAAACAAGCGCAACCCAGGACTTCTGGAGAGTTCAGTGGGATCCCCACCCCGCC 960  
 Qy 967 GGCTGAAAGAAAGAAAGAGCGTGACGCTGCTGGACGCTGGGCGACCCCTACTTACGCTG 1026  
 Db 961 GGCTGAAAGAAAGAAAGAGCGTGACGCTGCTGGACGCTGGGCGACCCCTACTTACGCTG 1020  
 Qy 1027 CCCTTGACGAGGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACACGAG 1086  
 Db 1021 CCCTTGACGAGGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACACGAG 1080  
 Qy 1087 ACCCGCGGATCCGCTACCAAGTACCAAGTCTGCTGCCCGAGGCTGGAGGCGAGCCCGAC 1146  
 Db 1081 ACCCGCGGATCCGCTACCAAGTACCAAGTCTGCTGCCCGAGGCTGGAGGCGAGCCCGAC 1140  
 Qy 1147 ATCTTCCAGAGCAGATGACCAAGATCTTGGAGCCCTTCCGCGCGCAACCCCGAGATC 1206  
 Db 1141 ATCTTCCAGAGCAGATGACCAAGATCTTGGAGCCCTTCCGCGCGCAACCCCGAGATC 1200  
 Qy 1207 GTGATCTACAGCCCGCTGTACGTGGCAGCAGCTGAGATCGGCGAGCACCAGCGCC 1266  
 Db 1201 GTGATCTACAGCCCGCTGTACGTGGCAGCAGCTGAGATCGGCGAGCACCAGCGCC 1260  
 Qy 1267 ARGATCGAGGAGTGGCGAGCAGCTTCTGCGTGGGCTTCAACACCCCGAGAGAG 1326  
 Db 1261 ARGATCGAGGAGTGGCGAGCAGCTTCTGCGTGGGCTTCAACACCCCGAGAGAG 1320  
 Qy 1327 CAGCAGAAGAGCCCGCTTCTGCGCATCGAGCTGACCCCGAGAGTGGACCGTGCGAG 1386  
 Db 1321 CAGCAGAAGAGCCCGCTTCTGCGCATCGAGCTGACCCCGAGAGTGGACCGTGCGAG 1380  
 Qy 1387 CCATCGAGTGGCGAGAGGAGAGTGGACCGTGAACGACATCCAGAGCTGGTGGCG 1446  
 Db 1381 CCATCGAGTGGCGAGAGGAGAGTGGACCGTGAACGACATCCAGAGCTGGTGGCG 1440  
 Qy 1447 AAGCTGAATGGCGCAGCAGATCTACCCCGCATCAAGTGGCGCGAGTGGCAAGCTG 1506  
 Db 1441 AAGCTGAATGGCGCAGCAGATCTACCCCGCATCAAGTGGCGCGAGTGGCAAGCTG 1500  
 Qy 1507 CTGCGGCGCGCAAGCGCTGACCGACATCTGCTCCCTGACCGAGAGCGCGAGTGGAG 1566  
 Db 1501 CTGCGGCGCGCAAGCGCTGACCGACATCTGCTCCCTGACCGAGAGCGCGAGTGGAG 1560  
 Qy 1567 CTGCGCGAGAACCGCGAGATCTGCGCGAGCGCTGACCGGCTGCTACGACCCCGAGC 1626  
 Db 1561 CTGCGCGAGAACCGCGAGATCTGCGCGAGCGCTGACCGGCTGCTACGACCCCGAGC 1620  
 Qy 1627 AAGGACTGTGCGCGAGATCCAGAGCAGGCGCACGACAGTGGACCTTACAGATCTAC 1686  
 Db 1621 AAGGACTGTGCGCGAGATCCAGAGCAGGCGCACGACAGTGGACCTTACAGATCTAC 1680  
 Qy 1687 CAGGAGCCCTTCAAGACCTGAGACCGGACAGTACGCAAGATCGCACCGCCCGACACC 1746  
 Db 1681 CAGGAGCCCTTCAAGACCTGAGACCGGACAGTACGCAAGATCGCACCGCCCGACACC 1740  
 Qy 1747 AACGAGTGAAGCAGTGAACCGAGGCGCTGACAGAGATCGCCATGGAGAGCATCTGATC 1806  
 Db 1741 AACGAGTGAAGCAGTGAACCGAGGCGCTGACAGAGATCGCCATGGAGAGCATCTGATC 1800  
 Qy 1807 TGGGCGAAGACCCCGAGTTCGCTCCCATCCAGAGAGACCTGGAGACCTGGTGG 1866  
 Db 1801 TGGGCGAAGACCCCGAGTTCGCTCCCATCCAGAGAGACCTGGAGACCTGGTGG 1860  
 Qy 1867 ACCGACTACTGGCAGGCGCACTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTG 1926  
 Db 1861 ACCGACTACTGGCAGGCGCACTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTG 1920  
 Qy 1927 GTGAAGCTGTGTTACCAAGCTGGAGAGAGCCCATCATCTGGCGCGAGACCTTCTACGTT 1986

Db 1921 GTGAAGCTGTGTTACCAAGCTGGAGAGAGCCCATCATCGCGCCGAGACCTTCTACGTT 1980  
 Qy 1987 GACGGCCCGCCCAACCCGCGAGACCAAGATCGGCAAGCGCGGTACGTGACCGACCGGGC 2046  
 Db 1981 GACGGCCCGCCCAACCCGCGAGACCAAGATCGGCAAGCGCGGTACGTGACCGACCGGGC 2040  
 Qy 2047 CGGCAAGAGATCTGAGCTTACCGGAGACCAACCAAGAGACCGAGTGGAGGCGCATC 2106  
 Db 2041 CGGCAAGAGATCTGAGCTTACCGGAGACCAACCAAGAGACCGAGTGGAGGCGCATC 2100  
 Qy 2107 CAGTGGCCCTCGAGGACAGCGCAGAGTGAACATCGTGACCGACAGCCAGTACGCC 2166  
 Db 2101 CAGTGGCCCTCGAGGACAGCGCAGAGTGAACATCGTGACCGACAGCCAGTACGCC 2160  
 Qy 2167 CTGGGCAATCATCGAGCCCGAGCCGACAGAGCGAGCGAGTGGTGAACAGATCATC 2226  
 Db 2161 CTGGGCAATCATCGAGCCCGAGCCGACAGAGCGAGCGAGTGGTGAACAGATCATC 2220  
 Qy 2227 GAGCAGCTGATCAAGAGAGAGAGTGTACTCAGCTGGGTGCCGCCCAAGAGGCGCATC 2286  
 Db 2221 GAGCAGCTGATCAAGAGAGAGAGTGTACTCAGCTGGGTGCCGCCCAAGAGGCGCATC 2280  
 Qy 2287 GGCGCAACGAGCAGATCGACAAGCTGGTGGAGCAAGGCGCATCCGCAAGGTGTGTTCTG 2346  
 Db 2281 GGCGCAACGAGCAGATCGACAAGCTGGTGGAGCAAGGCGCATCCGCAAGGTGTGTTCTG 2340  
 Qy 2347 GACGGCATCGATGGCGCATCGTGATCTACCATGACGATGAGCAGCAGCTGTACGTGGGCGC 2406  
 Db 2341 GACGGCATCGATGGCGCATCGTGATCTACCATGACGATGAGCAGCAGCTGTACGTGGGCGC 2400  
 Qy 2407 GGCGCCCTAGGATCGATTAAGCTTCCCGGGCTAGCACCGGT 2451  
 Db 2401 GGCGCCCTAGGATCGATTAAGCTTCCCGGGCTAGCACCGGT 2445

RESULT 3

ADCI3264  
 ID ADC13264 standard; DNA; 2445 BP.  
 XX AC ADC13264;  
 XX DT 18-DEC-2003 (first entry)  
 XX DE DNA of HIV construct p2Pol-opt-YMWM\_C SEQ ID NO 43.  
 XX KW expression cassette; HIV Gag; Env; Int; Nef; p15NaseH; Pol; Tat; Prot;  
 XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
 XX OS Human immunodeficiency virus.  
 XX PN WO2003004620-A2.  
 XX PD 16-JAN-2003.  
 XX PF 05-JUL-2002; 2002WO-US021420.  
 XX PR 05-JUL-2001; 2001US-0303192P.  
 XX PR 31-AUG-2001; 2001US-0316860P.  
 XX PR 16-JAN-2002; 2002US-0349871P.  
 XX PA (CHIR ) CHIRON CORP.  
 XX PA (UYST-) UNIV STELLENBOSCH.  
 XX PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
 XX DR WPI; 2003-221593/21.  
 XX PT New expression cassette comprising a polynucleotide sequence encoding a  
 PT polypeptide including an HIV Gag, Env, Int, Nef, p15NaseH, Pol, Tat,  
 PT Prot, or Rev polypeptide, useful for immunization, or generating  
 PT packaging cell lines.  
 PS Disclosure; Fig 40; 301pp; English.

XX The invention relates to a novel expression cassette comprising a  
 CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
 CC Int, Nef, p15NaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
 CC expression cassette can be used to treat HIV type C by gene therapy or  
 CC used in the development of a vaccine. The gene delivery vector is  
 CC administered intramuscularly, intravenously, intranasally,  
 CC subcutaneously, intradermally, transdermally, intravaginally,  
 CC intrarectally, orally or intravenously. The expression cassette is useful  
 CC for immunisation, generating packaging cell lines and producing HIV  
 CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
 CC Type C related sequence of the invention.  
 XX  
 SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 99.4%; Score 2443.4; DB 9; Length 2445;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-294;  
 Matches 2444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 GCCACATGGCGGAGCCATGAGCCAGCGCCACAGCGCCAAATCCTTGATCAGCGGAGC 66  
 DB 1 GCCACATGGCGGAGCCATGAGCCAGCGCCACAGCGCCAAATCCTTGATCAGCGGAGC 60

QY 67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGCCATC 126  
 DB 61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGCCATC 120

QY 127 GCCCGCAACTGCGCGCCCCCGCGAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCAC 186  
 DB 121 GCCCGCAACTGCGCGCCCCCGCGAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCAC 180

QY 187 CAGATGAAGGACTGACCGGAGCGCCAGCGCCAACTTCTTCCGAGGAGCTTGGCCCTTCCC 246  
 DB 181 CAGATGAAGGACTGACCGGAGCGCCAGCGCCAACTTCTTCCGAGGAGCTTGGCCCTTCCC 240

QY 247 CAGGCGAAGCGCGGAGTTCCTCCAGCGAGCAGACCGCGCCACAGCGCCACAGCGGC 306  
 DB 241 CAGGCGAAGCGCGGAGTTCCTCCAGCGAGCAGACCGCGCCACAGCGCCACAGCGGC 300

QY 307 GAGCTGCAAGGTGCGCGGGGCAAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCAC 366  
 DB 301 GAGCTGCAAGGTGCGCGGGGCAAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCAC 360

QY 367 CTGAATCTTCCCGCAGATCACCCTGTGGCAGCGCCCGCTGTGTAGCATCAAGGTGGCGGC 426  
 DB 361 CTGAATCTTCCCGCAGATCACCCTGTGGCAGCGCCCGCTGTGTAGCATCAAGGTGGCGGC 420

QY 427 CAGATCAAGAGGCGCCCTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAC 486  
 DB 421 CAGATCAAGAGGCGCCCTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAC 480

QY 487 CTGCGCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCG 546  
 DB 481 CTGCGCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCG 540

QY 547 CAGTACGACCATCTCTGATCGAGATCTGCGCGCAGAGAGGCATCGGCACCGTCTGATC 606  
 DB 541 CAGTACGACCATCTCTGATCGAGATCTGCGCGCAGAGAGGCATCGGCACCGTCTGATC 600

QY 607 GSCCCACCCCGGTGAACATCATCTGCGCGCGCAACATGCTGACCCAGCTGGGCTGACCCCTG 666  
 DB 601 GSCCCACCCCGGTGAACATCATCTGCGCGCGCAACATGCTGACCCAGCTGGGCTGACCCCTG 660

QY 667 AACTTCCCGCATCAGCCCATGAGACCGTGGCGCGTGAAGTGAAGTGAAGTGAAGTGAAGT 726  
 DB 661 AACTTCCCGCATCAGCCCATGAGACCGTGGCGCGTGAAGTGAAGTGAAGTGAAGTGAAGT 720

QY 727 CCCAAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 786  
 DB 721 CCCAAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 780

QY 787 GAGGAGATGGAGAGGAGGCAAGATCACCAGATCGGCGCCCGAGAACCCCTCAACACCC 846  
 DB 787 GAGGAGATGGAGAGGAGGCAAGATCACCAGATCGGCGCCCGAGAACCCCTCAACACCC 840

DB 781 GAGGAGATGGAGAGGAGGCAAGATCACCAGATCGGCGCCCGAGAACCCCTCAACACCC 840  
 QY 847 CCCTGTGTGGCCATCAAGAGAGGAGCAGCAGCAAGATGCGCAAGCTGTGTGATTCGCG 906  
 DB 841 CCCTGTGTGGCCATCAAGAGAGGAGCAGCAGCAAGATGCGCAAGCTGTGTGATTCGCG 900

QY 907 GAGCTGAACAGCGCCACCCAGGACTTCTGGAGGTGACGTGGCATCCCCACCCCGCC 966  
 DB 901 GAGCTGAACAGCGCCACCCAGGACTTCTGGAGGTGACGTGGCATCCCCACCCCGCC 960

QY 967 GGCTTGAAGAGAGAGAGCGCTGACCGTGTGGAGCTGGCGGAGCGCTTCTTCAGCGTG 1026  
 DB 961 GGCTTGAAGAGAGAGAGCGTGTGACCGTGTGGAGCTGGCGGAGCGCTTCTTCAGCGTG 1020

QY 1027 CCCTTGGAGGAGACTTCCGCAAGTACACCGGCTTCAACATCCCGAGCATCAACACAGAG 1086  
 DB 1021 CCCTTGGAGGAGACTTCCGCAAGTACACCGGCTTCAACATCCCGAGCATCAACACAGAG 1080

QY 1087 ACCCCCGGATCCGCTACCAAGTACACGCTGTGCCCGAGGCTGGAAGGCGAGCCCGCAG 1146  
 DB 1081 ACCCCCGGATCCGCTACCAAGTACACGCTGTGCCCGAGGCTGGAAGGCGAGCCCGCAG 1140

QY 1147 ATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206  
 DB 1141 ATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

QY 1207 GTGATCTACAGGCCCGCTGTACGTGGCGAGCGACTGTGGAGTGGCGCAGCACCGCGCC 1266  
 DB 1201 GTGATCTACAGGCCCGCTGTACGTGGCGAGCGACTGTGGAGTGGCGCAGCACCGCGCC 1260

QY 1267 AAGATCGAGAGCTGCGCAAGCACTCTGCTGCGCTGGGGTTTCAACACCCCGCAAGAGAG 1326  
 DB 1261 AAGATCGAGAGCTGCGCAAGCACTCTGCTGCGCTGGGGTTTCAACACCCCGCAAGAGAG 1320

QY 1327 CACGAGAGAGGCCCGCTTCTTCCCATCGAGCTGCACCCCGACAGTGGACCGTGGAG 1386  
 DB 1321 CACGAGAGAGGCCCGCTTCTTCCCATCGAGCTGCACCCCGACAGTGGACCGTGGAG 1380

QY 1387 CCATCGAGCTGCGCGAGAGGAGAGCTGGAACCGTGAACGACATCCAGAGAGCTGTGGGC 1446  
 DB 1381 CCATCGAGCTGCGCGAGAGGAGAGCTGGAACCGTGAACGACATCCAGAGAGCTGTGGGC 1440

QY 1447 AAGCTGAATGGGCGAGCGAGATCTACCCCGGATCAAGGTGGCGCGAGCTGTGCAAGCTG 1506  
 DB 1441 AAGCTGAATGGGCGAGCGAGATCTACCCCGGATCAAGGTGGCGCGAGCTGTGCAAGCTG 1500

QY 1507 CTGCGCGGCGCAAGGCCCTGACCGACATCTGCGCCCTGACCGAGAGGCGCGAGCTGGAG 1566  
 DB 1501 CTGCGCGGCGCAAGGCCCTGACCGACATCTGCGCCCTGACCGAGAGGCGCGAGCTGGAG 1560

QY 1567 CTGCGCGAGAACCGCGAGATCTTGGCGAGCCCGTGCACCGCGTGTATCTACGACCCCGAG 1626  
 DB 1561 CTGCGCGAGAACCGCGAGATCTTGGCGAGCCCGTGCACCGCGTGTATCTACGACCCCGAG 1620

QY 1627 AAGGACCTGTGGCGGAGATCCAGAGCAGGCGCCAGCAGCTGGACCTTACGAGATCTAC 1686  
 DB 1621 AAGGACCTGTGGCGGAGATCCAGAGCAGGCGCCAGCAGCTGGACCTTACGAGATCTAC 1680

QY 1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGCACCCCGCACACC 1746  
 DB 1681 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGCACCCCGCACACC 1740

QY 1747 AAGCAGTGAAGCAGCTGACCGAGCCGTGCGAGAGATCGCCATGCGAGAGCATCTGTGATC 1806  
 DB 1741 AAGCAGTGAAGCAGCTGACCGAGCCGTGCGAGAGATCGCCATGCGAGAGCATCTGTGATC 1800

QY 1807 TGGGGCAAGACCCCGAAGTTCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCT 1866  
 DB 1801 TGGGGCAAGACCCCGAAGTTCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCT 1860

QY 1867 ACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGAGTTCTGTGAACACCCCGCCCTG 1926  
 DB 1861 ACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGAGTTCTGTGAACACCCCGCCCTG 1920



QY 794 TGGAGAGGAGGCAAGATCACCAAGATCGCCCGGAGAACCCCTACAAACACCCCGTGT 853  
 Db 2267 TGGAGAGGAGGCAAGATCACCAAGATCGCCCGGAGAACCCCTACAAACACCCCGTGT 2326  
 QY 854 TCGCCATCAAGAAGAGAGCAGCAGCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGA 913  
 Db 2327 TCGCCATCAAGAAGAGAGCAGCAGCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGA 2386  
 QY 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCATCCCCACCCCGCGGCTGA 973  
 Db 2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCATCCCCACCCCGCGGCTGA 2446  
 QY 974 AGAAGAAGAGAGCGTGACCGTGTCTGGAGCGTGGCGCAGCCCTACTTCAGCGTGGCCCTGG 1033  
 Db 2447 AGAAGAAGAGAGCGTGACCGTGTCTGGAGCGTGGCGCAGCCCTACTTCAGCGTGGCCCTGG 2506  
 QY 1034 ACGAGGACTTCGGCAAGTACACCGGCTTCACCATCCCCAGCATCAACAACAGAGACCCCG 1093  
 Db 2507 ACGAGGACTTCGGCAAGTACACCGGCTTCACCATCCCCAGCATCAACAACAGAGACCCCG 2566  
 QY 1094 GCATCCGCTTACCAAGTACCAAGTGTCTGCCCCAGGGCTGGAAGGCGAGCCCCCAGCATCTTCC 1153  
 Db 2567 GCATCCGCTTACCAAGTGTCTGCCCCAGGGCTGGAAGGCGAGCCCCCAGCATCTTCC 2626  
 QY 1154 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213  
 Db 2627 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686  
 QY 1214 ACCAGGCCCTCTGTACGTGGGCGACGACCTGGAGATCGGCCAGCACCCGCGCAAGATCG 1273  
 Db 2687 ACCAGGCCCTCTGTACGTGGGCGACGACCTGGAGATCGGCCAGCACCCGCGCAAGATCG 2746  
 QY 1274 AGAGCTGCGCAAGACCTGTCTGCGTGGGGTTCAACACCCCGCAAGAGCAACGAGA 1333  
 Db 2747 AGAGCTGCGCAAGACCTGTCTGCGTGGGGTTCAACACCCCGCAAGAGCAACGAGA 2806  
 QY 1334 AGAGCCCCCTCTGTCCTGCGATCGAGCTGCACCCGCAAGTGGAGCGTGGAGCCCTCG 1393  
 Db 2807 AGAGCCCCCTCTGTCCTGCGATCGAGCTGCACCCGCAAGTGGAGCGTGGAGCCCTCG 2866  
 QY 1394 AGCTGCCGAGAGAGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGCGAAGCTGA 1453  
 Db 2867 AGCTGCCGAGAGAGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGCGAAGCTGA 2926  
 QY 1454 ACTGGGCGAGAGAGATCTACCCCGGACATCAAGTGGCGAGCTGGAGCTGGCGCG 1513  
 Db 2927 ACTGGGCGAGAGAGATCTACCCCGGACATCAAGTGGCGAGCTGGAGCTGGCGCG 2986  
 QY 1514 GCGCCAGGCGCTGACCGAGATCGTGCCTTGAACGAGAGCGCGAGCTGGAGCTGGCGCG 1573  
 Db 2987 GCGCCAGGCGCTGACCGAGATCGTGCCTTGAACGAGAGCGCGAGCTGGAGCTGGCGCG 3046  
 QY 1574 AGAACCGCGAGATCTCTGCGCGAGCCCGTGCACGCGGTGTACTACGACCCCGCAAGGAGC 1633  
 Db 3047 AGAACCGCGAGATCTCTGCGCGAGCCCGTGCACGCGGTGTACTACGACCCCGCAAGGAGC 3106  
 QY 1634 TGGTGGCCGAGATCCAGAGAGCGGGCCAGCAGCTGGACCTTACAGATCTTACAGGAGC 1693  
 Db 3107 TGGTGGCCGAGATCCAGAGAGCGGGCCAGCAGCTGGACCTTACAGATCTTACAGGAGC 3166  
 QY 1694 CTTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGGCACCGCCCGCACCAACGAGC 1753  
 Db 3167 CTTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGGCACCGCCCGCACCAACGAGC 3226  
 QY 1754 TGAAGCAGCTGACCGAGGCGGTGAGAGATGCCATGAGAGCATCTGTGATCTGGGGCA 1813  
 Db 3227 TGAAGCAGCTGACCGAGGCGGTGAGAGATGCCATGAGAGCATCTGTGATCTGGGGCA 3286  
 QY 1814 AGACCCCCAAGTTCCGCTGCGCATCCAGAAGAGACCTGGGAGACCTGGTGACCGACT 1873  
 Db 3287 AGACCCCCAAGTTCCGCTGCGCATCCAGAAGAGACCTGGGAGACCTGGTGACCGACT 3346

QY 1874 ACTGGCAGGCCACCTGGATCCCGGAGTGGGAGTCTGTGAACACCCCGCTGTGTGAAGC 1933  
 Db 3347 ACTGGCAGGCCACCTGGATCCCGGAGTGGGAGTCTGTGAACACCCCGCTGTGTGAAGC 3406  
 QY 1934 TGTGTACAGCTGGAGAGAGGCCCATCATCGGCGCGAGACCTTCTACGTGGAGCGGCG 1993  
 Db 3407 TGTGTACAGCTGGAGAGAGGCCCATCATCGGCGCGAGACCTTCTACGTGGAGCGGCG 3466  
 QY 1994 CGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGCTACGTGACCCGACCGGGGCCGAGCA 2053  
 Db 3467 CGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGCTACGTGACCCGACCGGGGCCGAGCA 3526  
 QY 2054 AGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTGACGGCCATCCAGCTGG 2113  
 Db 3527 AGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTGACGGCCATCCAGCTGG 3586  
 QY 2114 CCCTGAGAGACAGCGGACGAGGTGAACATCTGACCCGACCGAGTACGCCCTGGGCA 2173  
 Db 3587 CCCTGAGAGACAGCGGACGAGGTGAACATCTGACCCGACCGAGTACGCCCTGGGCA 3646  
 QY 2174 TCATCAGGCCACAGCCCGACAAGAGAGCGAGCTGTGTGAACCAAGATCATCGAGCAGC 2233  
 Db 3647 TCATCAGGCCACAGCCCGACAAGAGAGCGAGCTGTGTGAACCAAGATCATCGAGCAGC 3706  
 QY 2234 TGATCAAGAGAGAGAGGTGTACTGAGTGGTGCCTCCGCAAGGGCATCGGGCGCA 2293  
 Db 3707 TGATCAAGAGAGAGAGGTGTACTGAGTGGTGCCTCCGCAAGGGCATCGGGCGCA 3766  
 QY 2294 ACAGCAGATTCGAACAGCTGTGAGCAAGGGCATCCGCAAGGTGTCTTCTGGAGCGCA 2353  
 Db 3767 ACAGCAGATTCGAACAGCTGTGAGCAAGGGCATCCGCAAGGTGTCTTCTGGAGCGCA 3826  
 QY 2354 TCATCGCGCATCGTGTACTACAGTACATGAGACCGCTGTACGTGGGCGAGCGGCGCC 2413  
 Db 3827 TCATCGCGCATCGTGTACTACAGTACATGAGACCGCTGTACGTGGGCGAGCGGCGCC 3886  
 QY 2414 CTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451  
 Db 3887 CTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 3924

RESULT 5  
 ABL39960  
 ID ABL39960 standard; DNA; 2463 BP.  
 XX AC ABL39960;  
 XX DX 15-MAY-2002 (first entry)  
 XX DE Synthetic construct PR975YM SEQ ID NO:31.  
 XX KW Human immunodeficiency virus type C; antigenic HIV type C protein; nef;  
 XX KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;  
 XX KW immunostimulant; gene therapy; gene; ds.  
 XX OS Human immunodeficiency virus; type C.  
 XX OS Synthetic.  
 XX WO200204493-A2.  
 XX PD 17-JAN-2002.  
 XX PF 05-JUL-2001; 2001WO-US021241.  
 XX PF 05-JUL-2000; 2000US-00610313.  
 XX PR (CHIR ) CHIRON CORP.  
 XX PA (UYST-) UNIV STELLENBOSCH.  
 XX PI Zur Wegede J, Barnett SW, Engelbrecht S, Van Rensburg BJ;  
 XX DR WPI; 2002-154920/20.

PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in  
Pr applications including DNA immunization or generation of packaging cell  
XX lines, particularly in gene therapy.

XX Claim 1; Fig 9; 233pp; English.

CC The present invention describes expression cassettes comprising a  
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV  
CC type C polypeptides. The expression cassettes comprise any of the HIV  
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef  
CC (i). (i) have immunostimulant activity and can be used in gene therapy.  
CC The HIV type C polynucleotides are useful in applications including DNA  
CC immunisation, generation of packaging cell lines, and production of HIV  
CC Type C proteins. The polynucleotides are particularly useful in gene  
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
CC AEB06204 to AEB06215 represent sequences used in the exemplification of  
CC the present invention

XX SQ Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

Query Match 99.2%; Score 2436.2; DB 6; Length 2463;  
Best Local Similarity 99.6%; Pred. No. 2.6e-293;  
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

Qy	1	GTCAGCGCCACCATGCGCGAGCCATGAGCCAGCCACGAGCGCCACATCTCTGATGACG	60
Db	1	GTCAGCGCCACCATGCGCGAGCCATGAGCCAGCCACGAGCGCCACATCTCTGATGACG	60
Qy	61	CGCAGCACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGAGGGC	120
Db	61	CGCAGCACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGAGGGC	120
Qy	121	CACATGCGCCGCAACTGCGCGCCCGCCCGCGCAAGAGGGCTGTGGAAAGTGCGGCAAGGAG	180
Db	121	CACATGCGCCGCAACTGCGCGCCCGCCCGCGCAAGAGGGCTGTGGAAAGTGCGGCAAGGAG	180
Qy	181	GSCCAGCAGATGAGGACTGACAGCGCGCCAGCGCAACTTCTTCGCGAGGACCTGGCC	240
Db	181	GSCCAGCAGATGAGGACTGACAGCGCGCCAGCGCAACTTCTTCGCGAGGACCTGGCC	240
Qy	241	TTCCGCCAGGGCAAGGCGCGGAGTTCCAGCGAGCAGAAACCGCGCAACAGCCCCACC	300
Db	241	TTCCGCCAGGGCAAGGCGCGGAGTTCCAGCGAGCAGAAACCGCGCAACAGCCCCACC	300
Qy	301	AGCCGCGAGCTGCAGGTGCGGGGACAAACCCCGCAGCGAGCGCGCGCGCGCGCCAG	360
Db	301	AGCCGCGAGCTGCAGGTGCGGGGACAAACCCCGCAGCGAGCGCGCGCGCGCGCCAG	360
Qy	361	GGCACCCTGAACTTCCCGCCAGATCACCTGTGGCAGCGCCCGCTGTGAGCATCAAGGTG	420
Db	361	GGCACCCTGAACTTCCCGCCAGATCACCTGTGGCAGCGCCCGCTGTGAGCATCAAGGTG	420
Qy	421	GGCGCCAGATCAAGAGGCGCTGTGACACCGCGCGCGAGCAGACACCGTCTCGAGGAG	480
Db	421	GGCGCCAGATCAAGAGGCGCTGTGACACCGCGCGCGAGCAGACACCGTCTCGAGGAG	480
Qy	481	ATGAGCTCTCCCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAG	540
Db	481	ATGAGCTCTCCCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAG	540
Qy	541	GTGCGCCAGTACGACATCTGATCGAGATCTCGCGCAAGAGGCCATCGGCACCGTG	600
Db	541	GTGCGCCAGTACGACATCTGATCGAGATCTCGCGCAAGAGGCCATCGGCACCGTG	600
Qy	601	CTGATCGGCGCCACCCCGTGAAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGC	660
Db	601	CTGATCGGCGCCACCCCGTGAAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGC	660
Qy	661	ACCCTGAATTCCTCCATCAGCCCATTCGAGACCGTGCCTGTGAAGCTGAAGCCCGGATG	720
Db	661	ACCCTGAATTCCTCCATCAGCCCATTCGAGACCGTGCCTGTGAAGCTGAAGCCCGGATG	720
Qy	721	GACGGCCCCAAGGTGAGCAGTGGGCCCTTGACCGAGGAGAGATCAAGGGCCCTGACCGCC	780

Db	721	GACGGCCCCAAGGTGAGCAGTGGGCCCTTGACCGAGGAGAGATCAAGGGCCCTGACCGCC	780
Qy	781	ATCTCGAGGAGATGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTAC	840
Db	781	ATCTCGAGGAGATGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTAC	840
Qy	841	AACACCCCGCTGTTCCGCATCAAGAAGAAGACAGCACCAAGTGGCGCAAGCTGGTGAC	900
Db	841	AACACCCCGCTGTTCCGCATCAAGAAGAAGACAGCACCAAGTGGCGCAAGCTGGTGAC	900
Qy	901	TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGACAGCTGGGATCCCCAC	960
Db	901	TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGACAGCTGGGATCCCCAC	960
Qy	961	CCCGCGGCGCTGAAGAAGAAGAGCGTGTGAGCGTGTGAGCGTGGGCGAGCGCTTCTTC	1020
Db	961	CCCGCGGCGCTGAAGAAGAAGAGCGTGTGAGCGTGTGAGCGTGGGCGAGCGCTTCTTC	1020
Qy	1021	AGCGTGGCCCTGAGAGGACTTCCGCAAGTACACCGCTTCAACATCCCCAGCATCAAC	1080
Db	1021	AGCGTGGCCCTGAGAGGACTTCCGCAAGTACACCGCTTCAACATCCCCAGCATCAAC	1080
Qy	1081	AACGAGACCCCGGCTACCGCTACAGTACAACTGTCTGCCCGGAGGTGGAGGGGACG	1140
Db	1081	AACGAGACCCCGGCTACCGCTACAGTACAACTGTCTGCCCGGAGGTGGAGGGGACG	1140
Qy	1141	CCGACATCTTCCAGAGCAGCATGACCAAGTCTTGGAGCCCTTCCGCGCCCGCAACCCC	1200
Db	1141	CCGACATCTTCCAGAGCAGCATGACCAAGTCTTGGAGCCCTTCCGCGCCCGCAACCCC	1200
Qy	1201	GAGATCGTGTATACACGCGCCCTGTACGTGGGCGAGCGACCTGGAGATGGCGAGCAC	1260
Db	1201	GAGATCGTGTATACACGCGCCCTGTACGTGGGCGAGCGACCTGGAGATGGCGAGCAC	1260
Qy	1261	CCGCGCAAGATCGAGAGCTGCGCAAGCACTGTGTGCGTGGGGTTCACACCCCGGAC	1320
Db	1261	CCGCGCAAGATCGAGAGCTGCGCAAGCACTGTGTGCGTGGGGTTCACACCCCGGAC	1320
Qy	1321	AGAGACCCAGAGAGCGCCCTTCTGCGCAT-----CGAGCTCACCCCGACAG	1374
Db	1321	AGAGACCCAGAGAGCGCCCTTCTGCGCAT-----CGAGCTCACCCCGACAG	1374
Qy	1375	TGGAACCGTGCAGCCCATCGAGCTGCGCGAGAGAGCTGGAACCGTGAACGATCCAG	1434
Db	1375	TGGAACCGTGCAGCCCATCGAGCTGCGCGAGAGAGCTGGAACCGTGAACGATCCAG	1434
Qy	1435	AGCTGCTGGGCAAGCTGAACTGGGCGAGCGCAGATCTACCCCGCATCAAGTGGCGAG	1494
Db	1435	AGCTGCTGGGCAAGCTGAACTGGGCGAGCGCAGATCTACCCCGCATCAAGTGGCGAG	1494
Qy	1495	CTGTGCAAGCTGCTGCGCGCGCCCAAGGCGCTGTGACCGCATCTGTCCTTGAACGAG	1554
Db	1495	CTGTGCAAGCTGCTGCGCGCGCCCAAGGCGCTGTGACCGCATCTGTCCTTGAACGAG	1554
Qy	1555	GCGAGCTGGAGCTGGCGAGAACCGCGAGATCTCTGGGAGCGCGTGCAGCGCGGTGAC	1614
Db	1555	GCGAGCTGGAGCTGGCGAGAACCGCGAGATCTCTGGGAGCGCGTGCAGCGCGGTGAC	1614
Qy	1615	TACGACCCCGAGCAAGGACTGTGTGGCGAGATCCAGAAAGAGGGCCACGACGATGGAC	1674
Db	1615	TACGACCCCGAGCAAGGACTGTGTGGCGAGATCCAGAAAGAGGGCCACGACGATGGAC	1674
Qy	1621	TACGACCCCGAGCAAGGACTGTGTGGCGAGATCCAGAAAGAGGGCCACGACGATGGAC	1680
Db	1621	TACGACCCCGAGCAAGGACTGTGTGGCGAGATCCAGAAAGAGGGCCACGACGATGGAC	1680
Qy	1675	TACGACCTTACGAGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACGCCAGATGGC	1734
Db	1675	TACGACCTTACGAGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACGCCAGATGGC	1734
Qy	1735	ACCGCCCAACCAAGCGTGAAGCGCTGACCGAGGCGCTGACAGATCGCCATGGAG	1794
Db	1735	ACCGCCCAACCAAGCGTGAAGCGCTGACCGAGGCGCTGACAGATCGCCATGGAG	1794
Qy	1741	ACCGCCCAACCAAGCGTGAAGCGCTGACCGAGGCGCTGACAGATCGCCATGGAG	1800
Db	1741	ACCGCCCAACCAAGCGTGAAGCGCTGACCGAGGCGCTGACAGATCGCCATGGAG	1800
Qy	1795	AGCATCGTGTATCTGGGCGAGACCCCGAAGTTCGCGCTTCCCATCCAGAGGAGACCTGG	1854
Db	1795	AGCATCGTGTATCTGGGCGAGACCCCGAAGTTCGCGCTTCCCATCCAGAGGAGACCTGG	1854

Db 1801 AGCATCGTGTATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGAGAGACCTGG 1860  
 Qy 1855 GAGACTGTGTGACCACTACTGCGAGGCACTGAGTCCCGAGTGGAGTTGTTGAAAC 1914  
 Db 1861 GAGACTGTGTGACCACTACTGCGAGGCACTGAGTCCCGAGTGGAGTTGTTGAAAC 1920  
 Qy 1915 ACCCCCCCTGTGTGAGCTGTGTTACCACTGAGTGGAGAGGAGCCCATCATCGGCGCCGAG 1974  
 Db 1921 ACCCCCCCTGTGTGAGCTGTGTTACCACTGAGTGGAGAGGAGCCCATCATCGGCGCCGAG 1980  
 Qy 1975 ACCTTCTACGTGACCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTG 2034  
 Db 1981 ACCTTCTACGTGACCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTG 2040  
 Qy 2035 ACCGACCGGCGCGCGAGAGATCTGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2094  
 Db 2041 ACCGACCGGCGCGCGAGAGATCTGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100  
 Qy 2095 CTGAGGCGCATCCAGCTGCGCTGCGAGGACGCGGAGCGAGTGAACATCGTGACCGAC 2154  
 Db 2101 CTGAGGCGCATCCAGCTGCGCTGCGAGGACGCGGAGCGAGTGAACATCGTGACCGAC 2160  
 Qy 2155 AGCAGTACCGCTGCGCATCTCCAGGCGCCAGCGGACGAGCGAGCGAGTGTGTG 2214  
 Db 2161 AGCAGTACCGCTGCGCATCTCCAGGCGCCAGCGGACGAGCGAGCGAGTGTGTG 2220  
 Qy 2215 AACGAGATCATCGAGCAGCTGTCAAGAGAGGAGAGGTGTACTGAGCTGGTGTGCGGCC 2274  
 Db 2221 AACGAGATCATCGAGCAGCTGTCAAGAGAGGAGAGGTGTACTGAGCTGGTGTGCGGCC 2280  
 Qy 2275 CACAAGGCGATCGCGCGCAACGAGCAGATCGACAGCTGGTGTGAGCAAGGCGCATCGCAAG 2334  
 Db 2281 CACAAGGCGATCGCGCGCAACGAGCAGATCGACAGCTGGTGTGAGCAAGGCGCATCGCAAG 2340  
 Qy 2335 GTGCTGTTCTGGAACGCGATCGAGCGGCGCATCGTGTATCTACCACTACATGAGCAGACCTG 2394  
 Db 2341 GTGCTGTTCTGGAACGCGATCGAGCGGCGCATCGTGTATCTACCACTACATGAGCAGACCTG 2400  
 Qy 2395 TAGTGGGAGCGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2454  
 Db 2401 TAGTGGGAGCGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2460  
 Qy 2455 TTC 2457  
 Db 2461 TTC 2463

RESULT 6

ADCl3231  
 ID ADCl3231 standard; DNA; 3930 BP.  
 XX  
 AC ADCl3231;  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE DNA of HIV construct GagCompPolmutAtt\_C SEQ ID NO 10.  
 XX  
 KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
 KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
 XX  
 OS Human immunodeficiency virus.  
 XX  
 XX WO2003004620-A2.  
 XX  
 XX 16-JAN-2003.  
 XX  
 XX 05-JUL-2002; 2002WO-US021420.  
 XX  
 XX 05-JUL-2001; 2001US-0303192P.  
 XX  
 XX 31-AUG-2001; 2001US-0316860P.  
 XX  
 XX 16-JAN-2002; 2002US-0349871P.

PA (JYST-) UNIV STELLENBOSCH.  
 XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
 XX WPI; 2003-221593/21.  
 XX  
 XX New expression cassette comprising a polynucleotide sequence encoding a  
 XX polypeptide including an HIV gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
 XX Prot, or Rev polypeptide, useful for immunization, or generating  
 XX packaging cell lines.  
 XX  
 XX Disclosure; Fig 7; 301pp; English.  
 XX  
 XX The invention relates to a novel expression cassette comprising a  
 XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
 XX Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
 XX expression cassette can be used to treat HIV type C by gene therapy or  
 XX used in the development of a vaccine. The gene delivery vector is  
 XX administered intramuscularly, intramuscularly, intranasally,  
 XX subcutaneously, intradermally, transdermally, intravenously,  
 XX intrarectally, orally or intravenously. The expression cassette is useful  
 XX for immunisation, generating packaging cell lines and producing HIV  
 XX polypeptides. This polynucleotide sequence represents the DNA of an HIV  
 XX type C related sequence of the invention.  
 XX  
 XX Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;

Query Match 99.1%; Score 2434.8; DB 9; Length 3930;  
 Best Local Similarity 99.9%; Pred. No. 3.6e-293;  
 Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 14 TGCGGAGCGCATGAGCGCCACAGCGCCCAACATCTCTGATCAGCGGAGCAACTTCA 73  
 Db 1487 TGCGGAGCGCATGAGCGCCACAGCGCCCAACATCTCTGATCAGCGGAGCAACTTCA 1546  
 Qy 74 AGGGGCCCAAGCGCATCATCAAGTGTCTTCACTCGGCAAGAGGCGCCACATCGCCGCA 133  
 Db 1547 AGGGGCCCAAGCGCATCATCAAGTGTCTTCACTCGGCAAGAGGCGCCACATCGCCGCA 1606  
 Qy 134 ACTGCGCGCGCCCGCGCAAGAGGCGTGTGGAAGTGGCGCAAGAGGCGCCACAGATGA 193  
 Db 1607 ACTGCGCGCGCCCGCGCAAGAGGCGTGTGGAAGTGGCGCAAGAGGCGCCACAGATGA 1666  
 Qy 194 AGGACTGCAACGAGCGCGCCAGCGCCCAACTTCTTCCGAGGAGCTTCCGCTTCCCGAGGCA 253  
 Db 1667 AGGACTGCAACGAGCGCGCCAGCGCCCAACTTCTTCCGAGGAGCTTCCGCTTCCCGAGGCA 1726  
 Qy 254 AGGCGCGGAGTTCCTCCAGCGAGCGAGAAACCGCGCCCAACAGCCCGAGGAGTGC 313  
 Db 1727 AGGCGCGGAGTTCCTCCAGCGAGCGAGAAACCGCGCCCAACAGCCCGAGGAGTGC 1786  
 Qy 314 AGGTGCGGCGGCAAAACCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
 Db 1787 AGGTGCGGCGGCAAAACCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846  
 Qy 374 TCCCCCAGATCACCTGTGAGCGCGCCCTGTGTGAGCATCAAGTGTGGCGCGCGCATCA 433  
 Db 1847 TCCCCCAGATCACCTGTGAGCGCGCGCCCTGTGTGAGCATCAAGTGTGGCGCGCGCATCA 1906  
 Qy 434 AGGAGCGCGCTGTGGACACCG 493  
 Db 1907 AGGAGCGCGCTGTGGACACCG 1966  
 Qy 494 GCAAGTGGAGCG 553  
 Db 1967 GCAAGTGGAGCG 2026  
 Qy 554 ACCAGATCTGATCGAGATCTGCG 613  
 Db 2027 ACCAGATCTGATCGAGATCTGCG 2086





PR 16-JAN-2002; 2002US-0349871P.  
 XX (CHIR ) CHIRON CORP.  
 PA (UYST-) UNIV STELLENBOSCH.  
 XX  
 XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg BJ;  
 PI WPI; 2003-221593/21.  
 DR  
 XX  
 XX New expression cassette comprising a polynucleotide sequence encoding a  
 PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
 PT Prot, or Rev polypeptide, useful for immunization, or generating  
 PT packaging cell lines.  
 XX  
 XX Disclosure; Fig 8; 301pp; English.  
 PS  
 XX The invention relates to a novel expression cassette comprising a  
 CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
 CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
 CC expression cassette can be used to treat HIV type C by gene therapy or  
 CC used in the development of a vaccine. The gene delivery vector is  
 CC administered intramuscularly, intramuscularly, intranasally,  
 CC subcutaneously, intradermally, transdermally, intravaginally,  
 CC intrarectally, orally or intravenously. The expression cassette is useful  
 CC for immunisation, generating packaging cell lines and producing HIV  
 CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
 CC Type C related sequence of the invention.  
 XX  
 SQ Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;  
 Query March 99.1%; Score 2434.8; DB 9; Length 3930;  
 Best Local Similarity 99.9%; Pred. No 3 6e-293;  
 Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 14 TGCCGAGGCGATGAGCCAGGCGCACAGCGCCAAATCTGTAGCGGCGAGCACTTCA 73  
 DB 1487 TGCCGAGGCGATGAGCCAGGCGCACAGCGCCAAATCTGTAGCGGCGAGCACTTCA 1546  
 QY 74 AGGGCCCCAGCGCATCATCAAGTGTCTCACTGGCGGAAGGAGGCGCCATCGCCCGCA 133  
 DB 1547 AGGGCCCCAGCGCATCATCAAGTGTCTCACTGGCGGAAGGAGGCGCCATCGCCCGCA 1606  
 QY 134 ACTGCCGCGCCCCCGCAAGAGGGGCTGTGGAAGTGGCGCAAGGAGGCGCCACAGATGA 193  
 DB 1607 ACTGCCGCGCCCCCGCAAGAGGGGCTGTGGAAGTGGCGCAAGGAGGCGCCACAGATGA 1666  
 QY 194 AGGACTGACCGAGCGCGCGCAACTTCTTCGGGAGGACCTGGCTTCCCGCAGGGCA 253  
 DB 1667 AGGACTGACCGAGCGCGCGCAACTTCTTCGGGAGGACCTGGCTTCCCGCAGGGCA 1726  
 QY 254 AGGCGCGGAGTTCCTCCAGCGAGCAAGCCGCGCAAGCGCCCAAGCGCGCGAGCTGC 313  
 DB 1727 AGGCGCGGAGTTCCTCCAGCGAGCAAGCCGCGCAAGCGCCCAAGCGCGCGAGCTGC 1786  
 QY 314 AGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
 DB 1787 AGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846  
 QY 374 TCCCGCAGATCATCCCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433  
 DB 1847 TCCCGCAGATCATCCCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1906  
 QY 434 AGGAGGCGCTGTGAGCACCGCGCGCGCGAGCACACCGTGTGAGGAGATGAGCGTGC 493  
 DB 1907 AGGAGGCGCTGTGAGCACCGCGCGCGCGAGCACACCGTGTGAGGAGATGAGCGTGC 1966  
 QY 494 GCAAGTGAAGCGCGAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 553  
 DB 1967 GCAAGTGAAGCGCGAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2026  
 QY 554 ACCAGATCTGTGAGATCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 613

QY 614 CCCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTTGAATTC 673  
 DB 2087 CCCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTTGAATTC 2146  
 QY 674 CCATCAGCCCCATCGAGAGCGGTCGCGGTGAAGCTGAAGCCCGGCGATGAGCGGCGCGCAAG 733  
 DB 2147 CCATCAGCCCCATCGAGAGCGGTCGCGGTGAAGCTGAAGCCCGGCGATGAGCGGCGCGCAAG 2206  
 QY 734 TGAAGCAGTGGCCCCCTGACCGAGGAGAAGATCAAGCCCTTGACCGCCATCTTGGAGAGA 793  
 DB 2207 TGAAGCAGTGGCCCCCTGACCGAGGAGAAGATCAAGCCCTTGACCGCCATCTTGGAGAGA 2266  
 QY 794 TGAAGAGGAGGCGAAGATCAACCAAGATCGGCCCCCGAGAACCCCTACACACCCCGCTGT 853  
 DB 2267 TGAAGAGGAGGCGAAGATCAACCAAGATCGGCCCCCGAGAACCCCTACACACCCCGCTGT 2326  
 QY 854 TCGCCATCAAGAGAAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGGACTTCCCGGAGCTGA 913  
 DB 2327 TCGCCATCAAGAGAAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGGACTTCCCGGAGCTGA 2386  
 QY 914 ACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCAATCCCGCCAGCGCGCGCTGA 973  
 DB 2387 ACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCAATCCCGCCAGCGCGCGCTGA 2446  
 QY 974 AGAAGAGAGAGAGCGTACCGCTGTGAGCTGGCGGCGAGCGCTACTTTCAGCGTCCCGCTGG 1033  
 DB 2447 AGAAGAGAGAGAGCGTACCGCTGTGAGCTGGCGGCGAGCGCTACTTTCAGCGTCCCGCTGG 2506  
 QY 1034 ACAGAGGACTTCCGCAAGTACACCGCTTCCACATCCCGAGCATCAACAGAGACCCCG 1093  
 DB 2507 ACAGAGGACTTCCGCAAGTACACCGCTTCCACATCCCGAGCATCAACAGAGACCCCG 2566  
 QY 1094 GCATCCCTACGATCAACGCTGCTCCCGCAGGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1153  
 DB 2567 GCATCCCTACGATCAACGCTGCTCCCGCAGGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2626  
 QY 1154 AGAGCAGCATGACCAAGATCTCTGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCG 1213  
 DB 2627 AGAGCAGCATGACCAAGATCTCTGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCG 2686  
 QY 1214 ACAGAGGCGCGCTGTAGTGGGAGGAGCTGAGAGTGGCGAGGAGGAGGAGGAGGAGGAGGAG 1273  
 DB 2687 ACAGAGGCGCGCTGTAGTGGGAGGAGCTGAGAGTGGCGAGGAGGAGGAGGAGGAGGAGGAG 2746  
 QY 1274 AGGAGCTGCGCAAGCAGCTGCTGCGTGGGCGCTTCCACACCCCGCGCAAGAGCACCGAGA 1333  
 DB 2747 AGGAGCTGCGCAAGCAGCTGCTGCGTGGGCGCTTCCACACCCCGCGCAAGAGCACCGAGA 2806  
 QY 1334 AGGAGCGCGCTTCTGCGCGCATCGAGTGCACCCCGCAAGTGGAGCGGTGAGCGCGCGCGCG 1393  
 DB 2807 AGGAGCGCGCTTCTGCGCGCATCGAGTGCACCCCGCAAGTGGAGCGGTGAGCGCGCGCGCG 2866  
 QY 1394 AGCTGCGCGAGAGGAGGAGCTGGACCGTGAACGACATCCAGAGAGCTGGTGGCGAAGCTGA 1453  
 DB 2867 AGCTGCGCGAGAGGAGGAGCTGGACCGTGAACGACATCCAGAGAGCTGGTGGCGAAGCTGA 2926  
 QY 1454 ACTGGGCGCAGCAGATCTACCCCGCATCAAGTGGCGCGAGCTGTGCAAGCTGTGCGCG 1513  
 DB 2927 ACTGGGCGCAGCAGATCTACCCCGCATCAAGTGGCGCGAGCTGTGCAAGCTGTGCGCG 2986  
 QY 1514 CGGCGAAGGCGCTGACCGAGCATCGTGGCGCTGACCGAGGAGGCGCGAGCTGAGCGCGCG 1573  
 DB 2987 GCGCGAAGGCGCTGACCGAGCATCGTGGCGCTGACCGAGGAGGCGCGAGCTGAGCGCGCG 3046  
 QY 1574 AGAACCGCGAGATCTCTGCGGAGCGCGCTGACCGCGCTGTACTACGACCCCGAGAGGAG 1633  
 DB 3047 AGAACCGCGAGATCTCTGCGGAGCGCGCTGACCGCGCTGTACTACGACCCCGAGAGGAG 3106  
 QY 1634 TGCTGGCGGAGATCCAGAGCAGGCGCGCAGCAGCTGAGCTACCATCTTACCGAGGAG 1693

QY 1694 CTTTCAAGAACTGTAAGACCGGCAAGTACGCAAGATGCGACCGCCACACCAACGACG 1753  
 Db 3167 CTTTCAAGAACTGTAAGACCGGCAAGTACGCAAGATGCGACCGCCACACCAACGACG 3226  
 QY 1754 TGAAGCAGCTGACCGAGGCGGCGTGCAGAGATGCGCAGATGCGTATCTGCGGCA 1813  
 Db 3227 TGAAGCAGCTGACCGAGGCGGCGTGCAGAGATGCGCAGATGCGTATCTGCGGCA 3286  
 QY 1814 AGACCCCAAGTTCCGCTGCGCTCCATCCAGAAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873  
 Db 3287 AGACCCCAAGTTCCGCTGCGCTCCATCCAGAAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346  
 QY 1874 ACTGCGAGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGC 1933  
 Db 3347 ACTGCGAGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGC 3406  
 QY 1934 TGTGTACAGCTGAGAGAGGAGCCATCATCGCGCGGAGACCTTACGTGACAGCGCG 1993  
 Db 3407 TGTGTACAGCTGAGAGAGGAGCCATCATCGCGCGGAGACCTTACGTGACAGCGCG 3466  
 QY 1994 CGGCCAACCGGAGACCAAGATCGGCAAGGCGGCTAGTGAACCGACCGGGCCGGGAGA 2053  
 Db 3467 CGGCCAACCGGAGACCAAGATCGGCAAGGCGGCTAGTGAACCGACCGGGCCGGGAGA 3526  
 QY 2054 AGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGG 2113  
 Db 3527 AGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGG 3586  
 QY 2114 CCTGCGAGACAGCGGACGAGGTGAACATCGTGAACCGACAGCAGTACGCCCTGGGCA 2173  
 Db 3587 CCTGCGAGACAGCGGACGAGGTGAACATCGTGAACCGACAGCAGTACGCCCTGGGCA 3646  
 QY 2174 TCATCCAGGCCAGCGCCGACAGAGGAGAGCTGGTGAACCGACATCATCGAGCAGC 2233  
 Db 3647 TCATCCAGGCCAGCGCCGACAGAGGAGAGCTGGTGAACCGACATCATCGAGCAGC 3706  
 QY 2234 TGATCAAGAGAGAGAGTGTACTGTAGCTGGTGGTCCGCCCAAGAGGCAATCGCGGCA 2293  
 Db 3707 TGATCAAGAGAGAGAGTGTACTGTAGCTGGTGGTCCGCCCAAGAGGCAATCGCGGCA 3766  
 QY 2294 ACGACAGATCGACAAGCTGGTGAAGAGGAGCATCCGCAAGGTGTCTTCCTGGACCGCA 2353  
 Db 3767 ACGACAGATCGACAAGCTGGTGAAGAGGAGCATCCGCAAGGTGTCTTCCTGGACCGCA 3826  
 QY 2354 TCGATGGCGGATCGTGTATCTACAGTATACAGTACAGTGTGATGCGGACAGCGCGGCC 2413  
 Db 3827 TCGATGGCGGATCGTGTATCTACAGTATACAGTGTGATGCGGACAGCGCGGCC 3886  
 QY 2414 CTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2451  
 Db 3887 CTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 3924

RESULT 8  
 ID ACA03591  
 XX ACA03591 standard; DNA; 5184 BP.  
 AC ACA03591;  
 XX ACA03591;  
 DT 22-MAY-2003 (first entry)  
 XX  
 DE Synthetic DNA encoding immunogenic HIV peptide #74.  
 XX  
 KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
 KW gene therapy; packaging cell line; humoral immune response;  
 KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2003004657-A1.  
 XX  
 PD 16-JAN-2003.

PF 05-JUL-2002; 2002WO-US021421.  
 XX 05-JUL-2001; 2001US-0303192P.  
 PR 31-AUG-2001; 2001US-0316860P.  
 PR 16-JAN-2002; 2002US-0349728P.  
 PR 16-JAN-2002; 2002US-0349793P.  
 PR 16-JAN-2002; 2002US-0349871P.  
 XX (CHIR ) CHIRON CORP.  
 PA Zur Megede J, Barnett SW, Lian Y;  
 XX WPI; 2003-221602/21.  
 DR New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
 XX polypeptides, useful as immunogenic compositions or vaccines for  
 PT generating humoral or cellular immune responses against HIV in a subject,  
 PT especially humans.  
 XX Example 1; Fig 79; 262pp; English.  
 PS The invention describes a synthetic polynucleotide encoding 2 or more  
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
 CC derived from different HIV subtypes. The polynucleotide is useful for  
 CC immunisation, generation of packaging cell lines, or production of HIV  
 CC polypeptides. The polynucleotide and its encoded proteins are useful as  
 CC immunogenic compositions or vaccines for generating humoral or cellular  
 CC immune responses against HIV in a subject, or for inducing neutralising  
 CC antibodies against HIV. The gene delivery vector comprising the  
 CC polynucleotide is also useful for DNA immunisation of, or for generating  
 CC an immune response (e.g. a humoral or cellular immune response) in, a  
 CC subject such as a mammal, particularly a human. This sequence encodes a  
 CC human immunodeficiency virus immunogenic peptide  
 XX  
 SQ Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;

Query Match 99.1%; Score 2434.8; DB 7; Length 5184;  
 Best Local Similarity 99.9%; Pred. No. 3.4e-293;  
 Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 14 TGGCCGAGCCATGAGCCAGGCCACAGCGCCCAACATCTGTATGAGCGCAGCACTTCA 73  
 Db 2741 TGGCCGAGCCATGAGCCAGGCCACAGCGCCCAACATCTGTATGAGCGCAGCACTTCA 2800  
 QY 74 AGGGCCCCAAGCGCATCATCAAGTGTCTCAATGCGGCAAGAGGAGGCGCACATCGCCCCGA 133  
 Db 2801 AGGGCCCCAAGCGCATCATCAAGTGTCTCAATGCGGCAAGAGGAGGCGCACATCGCCCCGA 2860  
 QY 134 ACTGCGCGCCCCCGCCAGAGAGGGGTGCTGAAGTGGCGCAAGGAGGCGCCACAGATGA 193  
 Db 2861 ACTGCGCGCCCCCGCCAGAGAGGGGTGCTGAAGTGGCGCAAGGAGGCGCCACAGATGA 2920  
 QY 194 AGGACTGCACCGAGCGCCAGGCCAACCTTCTTCCGCGAGGACCTGGCCCTTCCCCAGGGCA 253  
 Db 2921 AGGACTGCACCGAGCGCCAGGCCAACCTTCTTCCGCGAGGACCTGGCCCTTCCCCAGGGCA 2980  
 QY 254 AGGCCCGCGAGTTCGCCAGCGAGCAGACCGCGCCCAAGCGCCCGCCAGCGCGAGCTGC 313  
 Db 2981 AGGCCCGCGAGTTCGCCAGCGAGCAGACCGCGCCCAAGCGCCCGCCAGCGCGAGCTGC 3040  
 QY 314 AGGTGCGCGCGCACAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
 Db 3041 AGGTGCGCGCGCACAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3100  
 QY 374 TCCCCCAGATCACCCTGTGTGCGAGCGCCCTTGTGTAGCATCAAGTGGCGCGCGCGATCA 433  
 Db 3101 TCCCCCAGATCACCCTGTGTGCGAGCGCCCTTGTGTAGCATCAAGTGGCGCGCGCGATCA 3160  
 QY 434 AGGAGGCCCTGTGTGAGCACCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493  
 Db 3161 AGGAGGCCCTGTGTGAGCACCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3220

Db 3221 GCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTATCAAGGTGCGCAGTACG 3280  
QY 554 ACCAGATCCTGATCGAGATCTCGGCAAGAAAGGCCATCGGCACCGTGTGATCGGCCCA 613  
Db 3281 ACCAGATCCTGATCGAGATCTCGGCAAGAAAGGCCATCGGCACCGTGTGATCGGCCCA 3340  
QY 614 CCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGCAACCTGAACCTCC 673  
Db 3341 CCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGCAACCTGAACCTCC 3400  
QY 674 CCATCAGGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGATGAGCGGCCCAAGG 733  
Db 3401 CCATCAGGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGATGAGCGGCCCAAGG 3460  
QY 734 TGAAGCAGTGGCCCTGACCGAGGAGAAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793  
Db 3461 TGAAGCAGTGGCCCTGACCGAGGAGAAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520  
QY 794 TGGAGAAAGAGGCAAGATCAACCAAGATCGGCCCGGAGAACCTTACAAACACCCCGTGT 853  
Db 3521 TGGAGAAAGAGGCAAGATCAACCAAGATCGGCCCGGAGAACCTTACAAACACCCCGTGT 3580  
QY 854 TCGCCATCAAGAAAGAGGAGACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGA 913  
Db 3581 TCGCCATCAAGAAAGAGGAGACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGA 3640  
QY 914 ACAAGCGCACCCAGGACTTCTGGAGGTGACAGTGGGCATCCCGACCCCGCGGCTGA 973  
Db 3641 ACAAGCGCACCCAGGACTTCTGGAGGTGACAGTGGGCATCCCGACCCCGCGGCTGA 3700  
QY 974 AGAAGAAAGAGCGTGCCTGTGAGCTGGGCGACGCTTCTACAGCTGGCCCTGG 1033  
Db 3701 AGAAGAAAGAGCGTGCCTGTGAGCTGGGCGACGCTTCTACAGCTGGCCCTGG 3760  
QY 1034 ACGAGGACTTCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAAGACCCCG 1093  
Db 3761 ACGAGGACTTCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAAGACCCCG 3820  
QY 1094 GCATCCGTTACAGTACACCTGTGCTGCCAGGGCTGGAAGGCGAGCCCAAGCATCTTCC 1153  
Db 3821 GCATCCGTTACAGTACACCTGTGCTGCCAGGGCTGGAAGGCGAGCCCAAGCATCTTCC 3880  
QY 1154 AGAGCAGCATGACCAAGATCTCGAGGCTTCCCGCGCCGCAACCCCGAGATCGTGTCT 1213  
Db 3881 AGAGCAGCATGACCAAGATCTCGAGGCTTCCCGCGCCGCAACCCCGAGATCGTGTCT 3940  
QY 1214 ACCAGGCCCTGTACGTGGGACGACCTGGAGATCGGCAGCACCGCGCCCAAGATCG 1273  
Db 3941 ACCAGGCCCTGTACGTGGGACGACCTGGAGATCGGCAGCACCGCGCCCAAGATCG 4000  
QY 1274 AGGAGCTGGCAAGCACTGTGCGCTGGGGTTCAACACCCCGCAAGAAAGACCAAGA 1333  
Db 4001 AGGAGCTGGCAAGCACTGTGCGCTGGGGTTCAACACCCCGCAAGAAAGACCAAGA 4060  
QY 1334 AGGAGCCCTTCTGCTGCTGAGCTGCAACCCGCAAGTGGACCGTGCAGGCCCATCG 1393  
Db 4061 AGGAGCCCTTCTGCTGCTGAGCTGCAACCCGCAAGTGGACCGTGCAGGCCCATCG 4120  
QY 1394 AGCTCCCGAGAGAGAGCTGGAACCGTGAACGACATCCAGAGCTGTGGGCAAGCTGA 1453  
Db 4121 AGCTCCCGAGAGAGAGCTGGAACCGTGAACGACATCCAGAGCTGTGGGCAAGCTGA 4180  
QY 1454 ACTGGGCGAGCAGATCTTACCCCGCATCAGGTGCGCAGCTGTGCAAGCTGTGCGCG 1513  
Db 4181 ACTGGGCGAGCAGATCTTACCCCGCATCAGGTGCGCAGCTGTGCAAGCTGTGCGCG 4240  
QY 1514 GCGCAAGGCCCTGACCGACATCGTGCCTTACCGAGGAGGCCGAGCTGGAGCTGGCGG 1573  
Db 4241 GCGCAAGGCCCTGACCGACATCGTGCCTTACCGAGGAGGCCGAGCTGGAGCTGGCGG 4300

Db 4301 AGAACCGGAGATCTTGGCGGAGCCCGTGCACGGCGTGTACTACGACCCCGCAGGACC 4360  
QY 1634 TGGTGGCCGAGATCCAGAAAGCAGGGCCACGACCAAGTGAACCTTACAGATCTTACAGGAGC 1693  
Db 4361 TGGTGGCCGAGATCCAGAAAGCAGGGCCACGACCAAGTGAACCTTACAGATCTTACAGGAGC 4420  
QY 1694 CTTTCAAGAACCTGAGACCGGCAAGTACGCAAGATGCGCACCGCCACCAACGAGC 1753  
Db 4421 CTTTCAAGAACCTGAGACCGGCAAGTACGCAAGATGCGCACCGCCACCAACGAGC 4480  
QY 1754 TGAAGCAGCTGACCGAGCGCTGCAAGATCGCCATGGAGAGCATCTGTGATCTGGGGCA 1813  
Db 4481 TGAAGCAGCTGACCGAGCGCTGCAAGATCGCCATGGAGAGCATCTGTGATCTGGGGCA 4540  
QY 1814 AGACCCCAAGTTCGCTGCTGCCATCCAGAAAGAGACCTGGGAGACCTGGTGGACCGACT 1873  
Db 4541 AGACCCCAAGTTCGCTGCTGCCATCCAGAAAGAGACCTGGGAGACCTGGTGGACCGACT 4600  
QY 1874 ACTGGCAGGCCACTGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGC 1933  
Db 4601 ACTGGCAGGCCACTGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGC 4660  
QY 1934 TGTGTTACAGCTGGAGAAAGGCCATCATCGGCGCGAGACCTTCTACGTGGAGCGG 1993  
Db 4661 TGTGTTACAGCTGGAGAAAGGCCATCATCGGCGCGAGACCTTCTACGTGGAGCGG 4720  
QY 1994 CCGCCAAACCGCAGACCAAGATCGGCAAGCGCGGCTTACGTGACCGACCGGGCGCGGCA 2053  
Db 4721 CCGCCAAACCGCAGACCAAGATCGGCAAGCGCGGCTTACGTGACCGACCGGGCGCGGCA 4780  
QY 2054 AGATCGTGGAGCTGACCGAGACCAACCAAGACCGAGCTGCAAGGCGCATTCAGCTGG 2113  
Db 4781 AGATCGTGGAGCTGACCGAGACCAACCAAGACCGAGCTGCAAGGCGCATTCAGCTGG 4840  
QY 2114 CCCTGAGAGCAGCGGCGAGGTTGAACATCGTGAACGACAGCCAGTACGCTGGGCA 2173  
Db 4841 CCCTGAGAGCAGCGGCGAGGTTGAACATCGTGAACGACAGCCAGTACGCTGGGCA 4900  
QY 2174 TCATCAGGCCCAAGCCGCAAGAGCGAGCGAGCTGTGTGAACCAAGATCATCGAGCAGC 2233  
Db 4901 TCATCAGGCCCAAGCCGCAAGAGCGAGCGAGCTGTGTGAACCAAGATCATCGAGCAGC 4960  
QY 2234 TGTATCAAGAGGAGAGGTGTACTTACCTGAGCTGGTGCCTCCGACAGGCGCATCGCGGCA 2293  
Db 4961 TGTATCAAGAGGAGAGGTGTACTTACCTGAGCTGGTGCCTCCGACAGGCGCATCGCGGCA 5020  
QY 2294 ACCAGCAGATCGCAAGAGCTGAGCAAGGCGATCGCAAGGCTGTGTTCCTGGAGCGGCA 2353  
Db 5021 ACCAGCAGATCGCAAGAGCTGAGCAAGGCGATCGCAAGGCTGTGTTCCTGGAGCGGCA 5080  
QY 2354 TCATCGGCGCATCGTGTATCTACCATGATAGAGACCTGTACGTGGGCGAGCGGCGG 2413  
Db 5081 TCATCGGCGCATCGTGTATCTACCATGATAGAGACCTGTACGTGGGCGAGCGGCGG 5140  
QY 2414 CTAGGATCGATTAAAGTTCCTCGGCGCTAGCACCGGT 2451  
Db 5141 CTAGGATCGATTAAAGTTCCTCGGCGCTAGCACCGGT 5178

RESULT 9

ADC13279

ID ADC13279 standard; DNA; 5184 BP.

XX ADC13279;

AC ADC13279;

DT 18-DEC-2003 (first entry)

DE DNA of HIV construct TatRevNefgagCpolina\_c seq ID NO 58.

XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;

KW



QY 1574 AGAACCGGAGATCCTGCGGAGCCGCTGACGCGGTGTACTACGACCCGAGCAAGGACC 1633  
Db 4301 AGAACCGGAGATCCTGCGGAGCCGCTGACGCGGTGTACTACGACCCGAGCAAGGACC 4360  
QY 1634 TGGTGGCCGAGATCCAGAAAGCAGGCGCCACGACGAGTGGACCTACCAATCTTACCAAGAGC 1693  
Db 4361 TGGTGGCCGAGATCCAGAAAGCAGGCGCCACGACGAGTGGACCTACCAATCTTACCAAGAGC 4420  
QY 1694 CTTTCAAGAACTGAAAGCCGCGCAAGTACGCAAGATGCGCACCGCCCAACCAACGACG 1753  
Db 4421 CTTTCAAGAACTGAAAGCCGCGCAAGTACGCAAGATGCGCACCGCCCAACCAACGACG 4480  
QY 1754 TGAAGCAGTACCGAGGCGCTGCAAGAGATGCGCATCGAGAGCATCTGTGATCTGGGGCA 1813  
Db 4481 TGAAGCAGTACCGAGGCGCTGCAAGAGATGCGCATCGAGAGCATCTGTGATCTGGGGCA 4540  
QY 1814 AGACCCCAAGTTCCGCTGCGCTGCCATCCAGAAAGGAGACCTGGGAGACCTGTGTGACCGACT 1873  
Db 4541 AGACCCCAAGTTCCGCTGCGCTGCCATCCAGAAAGGAGACCTGGGAGACCTGTGTGACCGACT 4600  
QY 1874 ACTGCAGGCACTGATGCCAGTGGAGTTGTTGAACACACCCCGCTGTGTGAGC 1933  
Db 4601 ACTGCAGGCACTGATGCCAGTGGAGTTGTTGAACACACCCCGCTGTGTGAGC 4660  
QY 1934 TGTGTACAGTGGAGAGGAGCCATCATCGGCGCGAGACCTTCTACGTGGACGCGC 1993  
Db 4661 TGTGTACAGTGGAGAGGAGCCATCATCGGCGCGAGACCTTCTACGTGGACGCGC 4720  
QY 1994 CGGCCAACCGGAGACCAAGATCGGAGAGGCGCGGTACGTGACCGACCGGGCGCGGAGA 2053  
Db 4721 CGGCCAACCGGAGACCAAGATCGGAGAGGCGCGGTACGTGACCGACCGGGCGCGGAGA 4780  
QY 2054 AGATCGTGAGCTGACCGAGACCAACCAAGAGACGAGTGGAGCCATTCAGCTGG 2113  
Db 4781 AGATCGTGAGCTGACCGAGACCAACCAAGAGACGAGTGGAGCCATTCAGCTGG 4840  
QY 2114 CCTCGAGACACCGGACGAGGTGAACATCGTACCGACAGCAGTACGCTTGGGCA 2173  
Db 4841 CCTCGAGACACCGGACGAGGTGAACATCGTACCGACAGCAGTACGCTTGGGCA 4900  
QY 2174 TCATCCAGCCAGCCGACGAGGAGGAGGAGTGGTGAACACAGATCATCGAGCAGC 2233  
Db 4901 TCATCCAGCCAGCCGACGAGGAGGAGGAGTGGTGAACACAGATCATCGAGCAGC 4960  
QY 2234 TGATCAAGAGGAGAGGTGTACTCTGAGTGGGTGCGCCGCCACCAAGGGCATTCGGCGCA 2293  
Db 4961 TGATCAAGAGGAGAGGTGTACTCTGAGTGGGTGCGCCGCCACCAAGGGCATTCGGCGCA 5020  
QY 2294 ACGAGCAGATCGACAGCTGGTGGACAGGCGATCGGAGGTGCTGTCTTGGAGCGCA 2353  
Db 5021 ACGAGCAGATCGACAGCTGGTGGACAGGCGATCGGAGGTGCTGTCTTGGAGCGCA 5080  
QY 2354 TCATGGCGCATCGTGTACTACCATGATGAGACCTGTACGTGGGCGAGCGCGGCC 2413  
Db 5081 TCATGGCGCATCGTGTACTACCATGATGAGACCTGTGTACGTGGGCGAGCGCGGCC 5140  
QY 2414 CTAGGATCGATTAAGCTTCCGCGGCTAGCACCGGT 2451  
Db 5141 CTAGGATCGATTAAGCTTCCGCGGCTAGCACCGGT 5178

RESULT 10

ACA03547  
ID ACA03547 standard; DNA; 2457 BP.

AC ACA03547;

XX 22-MAY-2003 (first entry)

DE Synthetic DNA encoding immunogenic HIV peptide #30.

XX Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;

KW gene therapy; packaging cell line; humoral immune response;  
XX cellular immune response; gene delivery vector; DNA immunisation; ds.  
XX Synthetic.

PN WO2003004657-A1.

XX 16-JAN-2003.

XX 05-JUL-2002; 2002WO-US021421.

XX 05-JUL-2001; 2001US-0303192P.

PR 31-AUG-2001; 2001US-0316860P.

PR 16-JAN-2002; 2002US-0349728P.

PR 16-JAN-2002; 2002US-0349793P.

PR 16-JAN-2002; 2002US-0349871P.

XX (CHIR ) CHIRON CORP.

XX Zur Megede J, Barnett SW, Lian Y;

XX WPI; 2003-221602/21.

XX New synthetic polynucleotides encoding antigenic HIV type B and/or type C

XX polypeptides, useful as immunogenic compositions or vaccines for

XX generating humoral or cellular immune responses against HIV in a subject,

XX especially humans.

XX Example 1; Fig 35; 262pp; English.

XX The invention describes a synthetic polynucleotide encoding 2 or more  
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
CC derived from different HIV subtypes. The polynucleotide is useful for  
CC immunisation, generation of packaging cell lines, or production of HIV  
CC polypeptides. The polynucleotide and its encoded proteins are useful as  
CC immunogenic compositions or vaccines for generating humoral or cellular  
CC immune responses against HIV in a subject, or for inducing neutralising  
CC antibodies against HIV. The gene delivery vector comprising the  
CC polynucleotide is also useful for DNA immunisation of, or for generating  
CC an immune response (e.g. a humoral or cellular immune response) in, a  
CC subject such as a mammal, particularly a human. This sequence encodes a  
CC human immunodeficiency virus immunogenic peptide

XX Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

XX Query Match 98.8%; Score 2428.6; DB 7; Length 2457;

XX Best Local Similarity 99.6%; Pred. No. 2.2e-292;

XX Matches 2447; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

QY 1 GTGACGCCACCATGGCGGAGGCGCATGAGCCAGGCCACCGCCCAACATCTGTGATGCGAG 60

Db 1 GTGACGCCACCATGGCGGAGGCGCATGAGCCAGGCCACCGCCCAACATCTGTGATGCGAG 60

QY 61 CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTCACTGCGGCAAGGAGGC 120

Db 61 CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTCACTGCGGCAAGGAGGC 120

QY 121 CACATCGCCCGCAACTGCGCGGCCCGCCCGCAAGAGGGCTGCTGGAAGTGCAGCAAGGAG 180

Db 121 CACATCGCCCGCAACTGCGCGGCCCGCCCGCAAGAGGGCTGCTGGAAGTGCAGCAAGGAG 180

QY 181 GGCCACCATGATGAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Db 181 GGCCACCATGATGAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

QY 241 TTCCCCCAGGCGAAGCGCCCGAGTTCCCGAGGAGAGAACCGCGCGCAACACGCCCCACC 300

Db 241 TTCCCCCAGGCGAAGCGCCCGAGTTCCCGAGGAGAGAACCGCGCGCAACACGCCCCACC 300

QY 301 AGCCGCGAGTGTGAGGTGCGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCAG 360

Db 301 AGCCGCGAGTGTGAGGTGCGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCAG 360



361 GGACCTGAACTTCCCTCCAGATCACTCTGTGACGCGCCCTCTGTGAGCATCAAGTG 420  
 361 GGACCTGAACTTCCCTCCAGATCACTCTGTGACGCGCCCTCTGTGAGCATCAAGTG 420  
 421 GGGGCGCAGATCAAGGAGGCGCTCTGACACCGGCGCGCGACACCGCTGTGAGGAG 480  
 421 GGGGCGCAGATCAAGGAGGCGCTCTGACACCGGCGCGCGACACCGCTGTGAGGAG 480  
 481 ATGAGCTGTCGGGCAAGTGAAGCCCAAGATGATGCGCGGATCGCGGCTTCATCAAG 540  
 481 ATGAGCTGTCGGGCAAGTGAAGCCCAAGATGATGCGCGGATCGCGGCTTCATCAAG 540  
 541 GTGCGCAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCAACCGTG 600  
 541 GTGCGCAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCAACCGTG 600  
 601 CTGATCGGCGCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
 601 CTGATCGGCGCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
 661 ACCCTGAATCTCCCATCAGCCCATCGACACGCTGCGCTGAAGCTGAAGCCCGCATG 720  
 661 ACCCTGAATCTCCCATCAGCCCATCGACACGCTGCGCTGAAGCTGAAGCCCGCATG 720  
 721 GACGCGCCCAAGTGAAGCAGTGCCTCTGACCGAGGAGAGATCAAGGCCCTGAACGCGC 780  
 721 GACGCGCCCAAGTGAAGCAGTGCCTCTGACCGAGGAGAGATCAAGGCCCTGAACGCGC 780  
 781 ATCTGCGAGAGATGAGAGAGGAGGCAAGATCAACAGATCGGCGCGGAGAGCCCTAC 840  
 781 ATCTGCGAGAGATGAGAGAGGAGGCAAGATCAACAGATCGGCGCGGAGAGCCCTAC 840  
 841 AACACCCCGCTGTTCGCAATCAAGAGAGAGAGAGATCAAGTGGCGCAAGCTGGTGGAC 900  
 841 AACACCCCGCTGTTCGCAATCAAGAGAGAGAGAGATCAAGTGGCGCAAGCTGGTGGAC 900  
 901 TTCGCGAGCTGAACAGAGGAGCCAGGACTTCTGAGAGTGAAGCTGGGCGATCCCCAC 960  
 901 TTCGCGAGCTGAACAGAGGAGCCAGGACTTCTGAGAGTGAAGCTGGGCGATCCCCAC 960  
 961 CCGCGCGCTTGAAG 1020  
 961 CCGCGCGCTTGAAG 1020  
 1021 AGCTGCGCTTGAAG 1080  
 1021 AGCTGCGCTTGAAG 1080  
 1081 AACGAG 1140  
 1081 AACGAG 1140  
 1141 CCGAG 1200  
 1141 CCGAG 1200  
 1201 GAGATCGATGATACAG 1260  
 1201 GAGATCGATGATACAG 1260  
 1261 CCGCGCAAGATCGAG 1320  
 1261 CCGCGCAAGATCGAG 1320  
 1321 AAG 1374  
 1321 AAG 1374  
 1375 TGGACCGTGCAGCCCATCGAGTGCCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434  
 1381 TGGACCGTGCAGCCCATCGAGTGCCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440

1441 AAGTGTGGCAGCTGAATGGGCGAGCCAGATCTACCCCGCATCAAGTGGCCAG 1500  
 1495 CTGTGCAAGCTGTGCGCGCGCAAGCCCTGACCGACATCTGTGCGCTTGAACGAGGAG 1554  
 1501 CTGTGCAAGCTGTGCGCGCGCAAGCCCTGACCGACATCTGTGCGCTTGAACGAGGAG 1560  
 1555 GCGGAGCTGGAGCTGGCGCGGAGAACCGCGAGATCTCTGCGAGCCCGTGACGCGCTGTAC 1614  
 1561 GCGGAGCTGGAGCTGGCGCGGAGAACCGCGAGATCTCTGCGAGCCCGTGACGCGCTGTAC 1620  
 1615 TACGACCCACCAAGAGACCTGTGTGGCGGAGATCCAGAAAGCAGAGGCGCAGACAGTGAAC 1674  
 1621 TACGACCCACCAAGAGACCTGTGTGGCGGAGATCCAGAAAGCAGAGGCGCAGACAGTGAAC 1680  
 1675 TACGAGATCTACGAGGAGCCCTTCAAGAACCTGAAAGCCGCGAGTACGCGCAAGATCGC 1734  
 1681 TACGAGATCTACGAGGAGCCCTTCAAGAACCTGAAAGCCGCGAGTACGCGCAAGATCGC 1740  
 1735 ACCGCGCAACCAAGAGAGCTGAAGCAGCTGACCGAGGCGCTGTGAGAGAGATCGCATGGAG 1794  
 1741 ACCGCGCAACCAAGAGAGCTGAAGCAGCTGACCGAGGCGCTGTGAGAGAGATCGCATGGAG 1800  
 1795 AGCATCTGATCTGGGCGAGACCCCGCAAGTTCGCGCTGCGCATCCAGAGAGAGACCTGG 1854  
 1801 AGCATCTGATCTGGGCGAGACCCCGCAAGTTCGCGCTGCGCATCCAGAGAGAGACCTGG 1860  
 1855 GAGACTCTGTGAGACCGCTACTGTGCGAGCCACCTGGATCCCGGAGTGGAGTTCGTGAAC 1914  
 1861 GAGACTCTGTGAGACCGCTACTGTGCGAGCCACCTGGATCCCGGAGTGGAGTTCGTGAAC 1920  
 1915 ACCCGCGCTGTGTGAAGCTGTGTACCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1974  
 1921 ACCCGCGCTGTGTGAAGCTGTGTACCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
 1975 ACCTTCTAGCTGAGCGCGCGCCAAACCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2034  
 1981 ACCTTCTAGCTGAGCGCGCGCCAAACCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040  
 2035 ACCGACCGGCGCGCGAGAGATCTGTGAGCTGTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAG 2094  
 2041 ACCGACCGGCGCGCGAGAGATCTGTGAGCTGTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100  
 2095 CTGCGAGGCGCATCCAGCTGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2154  
 2101 CTGCGAGGCGCATCCAGCTGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160  
 2155 AGCCAGTACGCGCTGGGCGATCATCCAGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2214  
 2161 AGCCAGTACGCGCTGGGCGATCATCCAGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
 2215 AACGAGATCATCGAGAGCTGATCAAG 2274  
 2221 AACGAGATCATCGAGAGCTGATCAAG 2280  
 2275 CACAAAGGCGCATCGCGCGCAACGAGCAGATCGACAGCTGGTGAAGAGAGAGAGAGAGAGAGAG 2334  
 2281 CACAAAGGCGCATCGCGCGCAACGAGCAGATCGACAGCTGGTGAAGAGAGAGAGAGAGAGAGAG 2340  
 2335 GTGCTGTCTTCTGAGCGGAGATCGATGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2394  
 2341 GTGCTGTCTTCTGAGCGGAGATCGATGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400  
 2395 TACGTGGGCGAGCGCGCGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451  
 2401 TACGTGGGCGAGCGCGCGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

RESULT 11  
 ADC13265  
 ID ADC13265 standard; DNA; 2457 BP.  
 XX



XX 18-DEC-2003 (first entry)  
 XX DNA of HIV construct p2Pol-opt-YM\_C SEQ ID NO 44.  
 DE expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
 KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
 XX Human immunodeficiency virus.  
 OS WO2003004620-A2.  
 XX PN 16-JAN-2003.  
 XX PD 05-JUL-2002; 2002WO-US021420.  
 XX PF 05-JUL-2001; 2001US-0303192P.  
 XX PR 31-AUG-2001; 2001US-0316860P.  
 XX PR 16-JAN-2002; 2002US-0349871P.  
 XX (CHIR) CHIRON CORP.  
 PA (UYST-) UNIV STELLENBOSCH.  
 XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
 PI WPI; 2003-221593/21.  
 XX New expression cassette comprising a polynucleotide sequence encoding a  
 PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
 PT Prot, or Rev polypeptide, useful for immunization, or generating  
 PT packaging cell lines.  
 XX Disclosure; Fig 41; 301pp; English.  
 XX The invention relates to a novel expression cassette comprising a  
 CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
 CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
 CC expression cassette can be used to treat HIV type C by gene therapy or  
 CC used in the development of a vaccine. The gene delivery vector is  
 CC administered intramuscularly, intramusosally, intranasally,  
 CC subcutaneously, intradermally, transdermally, intravaginally,  
 CC intrarectally, orally or intravenously. The expression cassette is useful  
 CC for immunisation, generating packaging cell lines and producing HIV  
 CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
 CC Type C related sequence of the invention.  
 XX Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;  
 SQ

Query Match 98.8%; Score 2428.6; DB 9; Length 2457;  
 Best Local Similarity 99.6%; Pred. No. 2.2e-292;  
 Matches 2447; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

QY 1 GTGAGCGCCATGCGCGAGCGCATGAGCCAGGCGCACAGCGCCAAATCTGATGCGAG 60  
 DB 1 GTGAGCGCCATGCGCGAGCGCATGAGCCAGGCGCACAGCGCCAAATCTGATGCGAG 60  
 QY 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCACTGCGGCAAGAGGGC 120  
 DB 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCACTGCGGCAAGAGGGC 120  
 QY 121 CACATGCGCCGCAATGCGCGCCCGCCCGCAAGAGAGGCTGTGGAAGTGGCGCAAGGAG 180  
 DB 121 CACATGCGCCGCAATGCGCGCCCGCCCGCAAGAGAGGCTGTGGAAGTGGCGCAAGGAG 180  
 QY 181 GSCCACCAGATGAAGGACTGACCGGCGCCAGGCGCACTTCTTCGCGAGGACTGTGCC 240  
 DB 181 GSCCACCAGATGAAGGACTGACCGGCGCCAGGCGCACTTCTTCGCGAGGACTGTGCC 240  
 QY 241 TTCTCCCGCAGGCAAGCGCGCGAGTTCCCGCAGCGAGCAACCGCGCCAAACAGCCCCACC 300  
 DB 241 TTCTCCCGCAGGCAAGCGCGCGAGTTCCCGCAGCGAGCAACCGCGCCAAACAGCCCCACC 300  
 QY 301 AGCGCGAGCTGAGGTGCGCGCGGACACACCCCGCAGCGAGGCGCGCGCGCGCGAG 360

DB 301 AGCGCGAGCTGAGGTGCGCGCGGACAAACCCCGCAGCGAGCGCGCGCGCGCGAG 360  
 QY 361 GGCACCCCTGAATTCCTCCCGCAGATCACCTGTGGCAGCGCCCTCTGTGAGCATCAAGGTG 420  
 DB 361 GGCACCCCTGAATTCCTCCCGCAGATCACCTGTGGCAGCGCCCTCTGTGAGCATCAAGGTG 420  
 QY 421 GCGCGCAGATCAAGGAGCCCTGTGTGGACAACCGCGCGCGAGCAGCACCGTGTGAGGAG 480  
 DB 421 GCGCGCAGATCAAGGAGCCCTGTGTGGACAACCGCGCGCGAGCAGCACCGTGTGAGGAG 480  
 QY 481 ATGAGCCTGCGCGCAAGTGGAGCCAGATGATCGCGCGCATCGCGGCTTCATCAAG 540  
 DB 481 ATGAGCCTGCGCGCAAGTGGAGCCAGATGATCGCGCGCATCGCGGCTTCATCAAG 540  
 QY 541 GTGCGCCAGTACGACGAGATCTCTGATCGAGATCTGTGGCAAGAAGCCATCGCACCGGTG 600  
 DB 541 GTGCGCCAGTACGACGAGATCTCTGATCGAGATCTGTGGCAAGAAGCCATCGCACCGGTG 600  
 QY 601 CTGATCGCGCCCAACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
 DB 601 CTGATCGCGCCCAACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
 QY 661 ACCTGAACTTCCCATCAGCCCATCGAGACCGTGCCTGGAAGTGAAGCCCGGCGATG 720  
 DB 661 ACCTGAACTTCCCATCAGCCCATCGAGACCGTGCCTGGAAGTGAAGCCCGGCGATG 720  
 QY 721 GACGCGCCCAAGTGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
 DB 721 GACGCGCCCAAGTGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
 QY 781 ATCTGCGAGGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840  
 DB 781 ATCTGCGAGGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840  
 QY 841 AACACCCCGTGTTCCTCATCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900  
 DB 841 AACACCCCGTGTTCCTCATCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900  
 QY 901 TTCCGCGAGCTGAACAGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960  
 DB 901 TTCCGCGAGCTGAACAGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960  
 QY 961 CCGCGCGGCTGAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020  
 DB 961 CCGCGCGGCTGAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020  
 QY 1021 AGCGTGCCTTGAACGAGGAGTTCGCAAGTACACCGCTTTCACCATCCCGAGCATCAAC 1080  
 DB 1021 AGCGTGCCTTGAACGAGGAGTTCGCAAGTACACCGCTTTCACCATCCCGAGCATCAAC 1080  
 QY 1081 AACGAGACCCCGGCTACCGCTACAGTACAGTGTCTGCCAGGCTGGAGGAGGAGGAG 1140  
 DB 1081 AACGAGACCCCGGCTACCGCTACAGTACAGTGTCTGCCAGGCTGGAGGAGGAGGAG 1140  
 QY 1141 CCGCAGATCTTCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCGCGCCCGCAGCC 1200  
 DB 1141 CCGCAGATCTTCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCGCGCCCGCAGCC 1200  
 QY 1201 GAGATGATGATTAACAGGCCCCCTGTGTACGTGGGAGGAGGAGGAGGAGGAGGAGGAG 1260  
 DB 1201 GAGATGATGATTAACAGGCCCCCTGTGTACGTGGGAGGAGGAGGAGGAGGAGGAGGAG 1260  
 QY 1261 CGCGCCAAAGATCGAGGAGTGGCAAGCCTGCTGCTGGGGCTTCAACCAACCCCGGAG 1320  
 DB 1261 CGCGCCAAAGATCGAGGAGTGGCAAGCCTGCTGCTGGGGCTTCAACCAACCCCGGAG 1320  
 QY 1321 AAGAAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1374  
 DB 1321 AAGAAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380  
 QY 1375 TGGACCGTGCAGGCGGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1434

Db 1381 TGGACCTGTCAGCCATCGAGCTGCCGAGAGGAGCTGGACCTGTAACGACATCCAG 1440  
 Qy 1435 AAGCTGTGGGCGAAGCTGAATCTGGGCCAGCCAGATCTAATCCCGGCAATCAAGGTGCGCCAG 1494  
 Db 1441 AAGCTGTGGGCGAAGCTGAATCTGGGCCAGCCAGATCTAATCCCGGCAATCAAGGTGCGCCAG 1500  
 Qy 1495 CTGTGAAGCTGTGTCGGCGCCAGCCCTGACGACATCTGTCGCCCTGACCGAGGAG 1554  
 Db 1501 CTGTGAAGCTGTGTCGGCGCCAGCCCTGACGACATCTGTCGCCCTGACCGAGGAG 1560  
 Qy 1555 GCGAGCTGAGCTGGCGCGGAGAACCGCGAGATCTGCGGAGCCCTGACCGCGGTGTAC 1614  
 Db 1561 GCGAGCTGAGCTGGCGCGGAGAACCGCGAGATCTGCGGAGCCCTGACCGCGGTGTAC 1620  
 Qy 1615 TACGACCCAGCAGGACCTGTGTCGGCGGAGATCCAGAGCAGGCGCACGACGATGTGACC 1674  
 Db 1621 TACGACCCAGCAGGACCTGTGTCGGCGGAGATCCAGAGCAGGCGCACGACGATGTGACC 1680  
 Qy 1675 TACGAGATCTACGAGGACCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGCGC 1734  
 Db 1681 TACGAGATCTACGAGGACCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGCGC 1740  
 Qy 1735 ACCGCCACACCAACGACGCTGACGAGCTGACGAGGCGCTGCGAGAGATCGCATGGAG 1794  
 Db 1741 ACCGCCACACCAACGACGCTGACGAGCTGACGAGGCGCTGCGAGAGATCGCATGGAG 1800  
 Qy 1795 AGCATCTGTGATCTGGGCGAAGACCCCAAGTTCGCGCTGCCCATCCAGAGGAGACCTGG 1854  
 Db 1801 AGCATCTGTGATCTGGGCGAAGACCCCAAGTTCGCGCTGCCCATCCAGAGGAGACCTGG 1860  
 Qy 1855 GAGACCTGTGGACCGGACTACTGCGAGCGCCACTGATCCCGAGTGGGAGTTCGTGAAC 1914  
 Db 1861 GAGACCTGTGGACCGGACTACTGCGAGCGCCACTGATCCCGAGTGGGAGTTCGTGAAC 1920  
 Qy 1915 ACCGCCCCCTGTGTGTAAGCTGTGTACGAGTGGAGAGGAGCCCATCATCGCGCGCGAG 1974  
 Db 1921 ACCGCCCCCTGTGTGTAAGCTGTGTACGAGTGGAGAGGAGCCCATCATCGCGCGCGAG 1980  
 Qy 1975 ACCTTCTACGTGAGCGGCGCCCAACCGGAGACCAAGATCGCAAGCGCGCTAGTGG 2034  
 Db 1981 ACCTTCTACGTGAGCGGCGCCCAACCGGAGACCAAGATCGCAAGCGCGCTAGTGG 2040  
 Qy 2035 ACCGACCGGCGCGGCGAAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2094  
 Db 2041 ACCGACCGGCGCGGCGAAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100  
 Qy 2095 CTGACGCGCATTCAGCTGGCCCTGACGAGCAGCGGCGAGGAGTGAACATCGTGACCGAC 2154  
 Db 2101 CTGACGCGCATTCAGCTGGCCCTGACGAGCAGCGGCGAGGAGTGAACATCGTGACCGAC 2160  
 Qy 2155 AGCCAGTACGCCCTGGGCTCATCCAGCGCCAGCCGACAGAGCGAGAGCGAGCTGGTG 2214  
 Db 2161 AGCCAGTACGCCCTGGGCTCATCCAGCGCCAGCCGACAGAGCGAGAGCGAGCTGGTG 2220  
 Qy 2215 AACGAGATCTACGAGCAGCTGATCAAGAGAGAGGTGTACTGAGCTGGTGCCCGCC 2274  
 Db 2221 AACGAGATCTACGAGCAGCTGATCAAGAGAGAGGTGTACTGAGCTGGTGCCCGCC 2280  
 Qy 2275 CACAGGCGATCGGCGGCAAGCAGATCGACAGCTGCTGAGCAGGAGGCGATCCGCGAG 2334  
 Db 2281 CACAGGCGATCGGCGGCAAGCAGATCGACAGCTGCTGAGCAGGAGGCGATCCGCGAG 2340  
 Qy 2335 GTGCTGTCTCTGAGCGGATCGATGGCGGATCGTGATCTACAGTACATGAGCAGCTG 2394  
 Db 2341 GTGCTGTCTCTGAGCGGATCGATGGCGGATCGTGATCTACAGTACATGAGCAGCTG 2400  
 Qy 2395 TACGTGGGCGAGCGGCGCTTAGATGATTAAGCTTCCCGGGGTAGCACCGGT 2451  
 Db 2401 TACGTGGGCGAGCGGCGCTTAGATGATTAAGCTTCCCGGGGTAGCACCGGT 2457

RESULT 12  
 ABL39959

ID ABL39959 standard; DNA; 2469 BP.  
 XX  
 AC ABL39959;  
 XX  
 DT 15-MAY-2002 (first entry)  
 XX  
 DE Synthetic construct PR975(+) SEQ ID NO:30.  
 XX  
 DE Human immunodeficiency virus type C; antigenic HIV type C protein;  
 KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;  
 KW immunostimulant; gene therapy; gene; ds.  
 XX  
 OS Human immunodeficiency virus; type C.  
 OS Synthetic.  
 XX  
 EN WO200204493-A2.  
 XX  
 PD 17-JAN-2002.  
 XX  
 PF 05-JUL-2001; 2001WO-US021241.  
 XX  
 PR 05-JUL-2000; 2000US-00610313.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (UIST-) UNIV STELLENBOSCH.  
 PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;  
 XX WPI; 2002-154920/20.  
 XX  
 PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in  
 PT applications including DNA immunization or generation of packaging cell  
 PT lines, particularly in gene therapy.  
 XX  
 PS Claim 1; Fig 8; 233pp; English.  
 XX  
 CC The present invention describes expression cassettes comprising a  
 CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV  
 CC type C polypeptides. The expression cassettes comprise any of the HIV  
 CC type C sequences encoding gag, pol, vif, vpr, tat, rev, vpu, env or nef  
 CC (i). (i) have immunostimulant activity and can be used in gene therapy.  
 CC The HIV type C polynucleotides are useful in applications including DNA  
 CC immunisation, generation of packaging cell lines, and production of HIV  
 CC type C proteins. The polynucleotides are particularly useful in gene  
 CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
 CC ABB06204 to ABB06215 represent sequences used in the exemplification of  
 CC the present invention  
 XX  
 SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;  
 Query Match 98.3%; Score 2415.4; DB 6; Length 2469;  
 Best Local Similarity 99.3%; Pred. No. 9.6e-291;  
 Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;  
 Qy 1 GTGCGAGCCACCATGGCGAGGCGCATGAGCCAGCCACGAGCCCAACATCCTGATGCGAG 60  
 Db 1 GTGCGAGCCACCATGGCGAGGCGCATGAGCCAGCCACGAGCCCAACATCCTGATGCGAG 60  
 Qy 61 GCGCAGCAATTCAGAGGCGCCCAAGCGATCATCAAGTGTCTTCACTGCGGAGAGGCGC 120  
 Db 61 GCGCAGCAATTCAGAGGCGCCCAAGCGATCATCAAGTGTCTTCACTGCGGAGAGGCGC 120  
 Qy 121 CACATCGCCCGCAACTGCGCGCCCGCCCGCGCAAGAGGCGTCTGGAAGTGGCGCAAGAG 180  
 Db 121 CACATCGCCCGCAACTGCGCGCCCGCCCGCGCAAGAGGCGTCTGGAAGTGGCGCAAGAG 180  
 Qy 181 GCGCAGCAGATGAAGGACTGACCGAGCGCCAGCCCACTTCTTCCCGAGGACTGGCC 240  
 Db 181 GCGCAGCAGATGAAGGACTGACCGAGCGCCAGCCCACTTCTTCCCGAGGACTGGCC 240  
 Qy 241 TTCCCCCGAGGCGCAAGGCGCGGAGTTCCCCAGCGAGCAGAACCGCGCAACAGCCCCACC 300  
 Db 241 TTCCCCCGAGGCGCAAGGCGCGGAGTTCCCCAGCGAGCAGAACCGCGCAACAGCCCCACC 300

QY 301 AGCGGAGCTGCAGGTGGCGGCGACAAACCCCGCAGGAGCGCGCGCGAGGGCCAG 360  
DB |||||  
QY 301 AGCGGAGCTGCAGGTGGCGGCGACAAACCCCGCAGGAGCGCGCGCGAGGGCCAG 360  
DB |||||  
QY 361 GGCACCTGAACTTCCCCCAGATCACCTGTGGCAGCGCCCTCGTGAGCATCAAGGTG 420  
DB |||||  
QY 361 GGCACCTGAACTTCCCCCAGATCACCTGTGGCAGCGCCCTCGTGAGCATCAAGGTG 420  
DB |||||  
QY 421 GCGGCGCAGATCAAGGAGCCCTGTGACACCGCGCGCGAGCACCCTGCTGAGGAG 480  
DB |||||  
QY 421 GCGGCGCAGATCAAGGAGCCCTGTGACACCGCGCGCGAGCACCCTGCTGAGGAG 480  
DB |||||  
QY 481 ATGAGCTGCCCGGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTATCAAG 540  
DB |||||  
QY 481 ATGAGCTGCCCGGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTATCAAG 540  
DB |||||  
QY 541 GTGGCGCAGTACGACAGATCTGTATCGAGATCTGGGCAAGAGGCGCATCGGCACGTG 600  
DB |||||  
QY 541 GTGGCGCAGTACGACAGATCTGTATCGAGATCTGGGCAAGAGGCGCATCGGCACGTG 600  
DB |||||  
QY 601 CTGATCGGCGCCACCCCGTGAAATCATCGGCGCGCAATGCTGACCCAGCTGGGCTGC 660  
DB |||||  
QY 601 CTGATCGGCGCCACCCCGTGAAATCATCGGCGCGCAATGCTGACCCAGCTGGGCTGC 660  
DB |||||  
QY 661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGTGAAGCGCGCATG 720  
DB |||||  
QY 721 GAGCGCCCAAGGTGAAGCTGGCCCTGACCGGAGGAGATCAAGGCGCTGACCGCC 780  
DB |||||  
QY 721 GAGCGCCCAAGGTGAAGCTGGCCCTGACCGGAGGAGATCAAGGCGCTGACCGCC 780  
DB |||||  
QY 781 ATCTGCGAGGATGAGAGGAGGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTAC 840  
DB |||||  
QY 841 AACACCCCGTGTGGCCATCAAGAGAGGACAGACCAAGTGGCGAGCTGTGAC 900  
DB |||||  
QY 841 AACACCCCGTGTGGCCATCAAGAGAGGACAGACCAAGTGGCGAGCTGTGAC 900  
DB |||||  
QY 901 TTCCGCGAGCTGAAACAGGCGCACCGAGCTTCTGGAGGTGCAGCTGGGATCCCGCAC 960  
DB |||||  
QY 901 TTCCGCGAGCTGAAACAGGCGCACCGAGCTTCTGGAGGTGCAGCTGGGATCCCGCAC 960  
DB |||||  
QY 961 CCGCGCGCTGAAAGAGAGAGCGTGAACGCTGTGAGCGTGGGCGAGCGCTTACTTC 1020  
DB |||||  
QY 961 CCGCGCGCTGAAAGAGAGAGCGTGAACGCTGTGAGCGTGGGCGAGCGCTTACTTC 1020  
DB |||||  
QY 1021 AGCGTCCCTGAGAGGAGCTTCCGCAAGTACCGCGCTTCAACATCCCGAGCATCAAC 1080  
DB |||||  
QY 1021 AGCGTCCCTGAGAGGAGCTTCCGCAAGTACCGCGCTTCAACATCCCGAGCATCAAC 1080  
DB |||||  
QY 1081 AACGAGACCCCGGATCCGCTACAGTACAACGCTGTGCCCGAGGCTGGAAGGCGAGC 1140  
DB |||||  
QY 1081 AACGAGACCCCGGATCCGCTACAGTACAACGCTGTGCCCGAGGCTGGAAGGCGAGC 1140  
DB |||||  
QY 1141 CCGAGATCTTCAGAGAGAGTGAACCAAGATCTTGAGAGCCCTTCCGCGCGCGCAACCC 1200  
DB |||||  
QY 1141 CCGAGATCTTCAGAGAGAGTGAACCAAGATCTTGAGAGCCCTTCCGCGCGCGCAACCC 1200  
DB |||||  
QY 1201 GAGATCGTATCTTACCA-----GGCCCGCTGTAGTGGGCGAGCAGCTGAGATCGGC 1254  
DB |||||  
QY 1201 GAGATCGTATCTTACAGTACATGAGACCTGTAGTGGGCGAGCAGCTGAGATCGGC 1260  
DB |||||  
QY 1255 CAGCACCGGCGCAGATCGAGAGCTGCGCAAGCAGCTGTGCTGGGCGCTTACACACC 1314  
DB |||||  
QY 1261 CAGCACCGGCGCAGATCGAGAGCTGCGCAAGCAGCTGTGCTGGGCGCTTACACACC 1320  
DB |||||  
QY 1315 CCGGCAAGAGACCAAGAGAGCGCCCTTCTGCCCCAT-----CGAGCTGACCGC 1368  
DB |||||  
QY 1321 CCGGCAAGAGACCAAGAGAGCGCCCTTCTGCCCCAT-----CGAGCTGACCGC 1380  
DB |||||

QY 1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGCTGGAACCGAC 1428  
DB |||||  
QY 1381 GACRAGTGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGCTGGAACCGAC 1440  
DB |||||  
QY 1429 ATCCAGAAAGCTGTGGCAAGCTGAATCTGGCGCAGCAGATCTACCCCGCATCAAGGTG 1488  
DB |||||  
QY 1441 ATCCAGAAAGCTGTGGCAAGCTGAATCTGGCGCAGCAGATCTACCCCGCATCAAGGTG 1500  
DB |||||  
QY 1489 CGCCAGCTGTGAAGCTGTGGCGCGCAAGGCCCTGACCGACATCTGCGCCCTGACC 1548  
DB |||||  
QY 1501 CGCCAGCTGTGAAGCTGTGGCGCGCAAGGCCCTGACCGACATCTGCGCCCTGACC 1560  
DB |||||  
QY 1549 GAGAGGCGCAGCTGTGAGCTGGCGGAGAACCGCGAGATCTGCGCGAGCCCGTGCACGGC 1608  
DB |||||  
QY 1561 GAGAGGCGCAGCTGTGAGCTGGCGGAGAACCGCGAGATCTGCGCGAGCCCGTGCACGGC 1620  
DB |||||  
QY 1609 GTCTACTACACCCAGCAGGAGACCTGGTGGCGGAGATCCAGAGCAGGCGCACGACCAAG 1668  
DB |||||  
QY 1621 GTCTACTACACCCAGCAGGAGACCTGGTGGCGGAGATCCAGAGCAGGCGCACGACCAAG 1680  
DB |||||  
QY 1669 TGGACCTTACCAGATCTACAGGAGCGCCCTTCAAGAACCTGAGAACCGGCAAGTACGCCAAG 1728  
DB |||||  
QY 1681 TGGACCTTACCAGATCTACAGGAGCGCCCTTCAAGAACCGGCAAGTACGCCAAG 1740  
DB |||||  
QY 1729 ATGGGACCCGCCACACCAAGAGCTGAGAGCAGCTGACCGAGGCCCTGACAGAGATCGCC 1788  
DB |||||  
QY 1741 ATGGCAGCCGCCACACCAAGAGCTGAGAGCAGCTGACCGAGGCCCTGACAGAGATCGCC 1800  
DB |||||  
QY 1789 ATGGAGAGCATCTGTGATCTGGGCGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAG 1848  
DB |||||  
QY 1801 ATGGAGAGCATCTGTGATCTGGGCGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAG 1860  
DB |||||  
QY 1849 ACTGGAGACCTGTGTGAGCCGACTGTGCGAGGCCACCTGAGATCCCGAGTGGGAGTTC 1908  
DB |||||  
QY 1861 ACTGGAGACCTGTGTGAGCCGACTGTGCGAGGCCACCTGATCCCGAGTGGGAGTTC 1920  
DB |||||  
QY 1909 GTGAACACCCCGCTGTGTGAGCTGTGATCCAGCTGAGAGGAGGCCATCATCGGC 1968  
DB |||||  
QY 1921 GTGAACACCCCGCTGTGTGAGCTGTGATCCAGCTGAGAGGAGGCCATCATCGGC 1980  
DB |||||  
QY 1969 GCGGAGACCTTCTACGTGGAGCGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGCGGC 2028  
DB |||||  
QY 1981 GCGGAGACCTTCTACGTGGAGCGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGCGGC 2040  
DB |||||  
QY 2029 TACGTGACCGACCGGGCGCGGAGAGCTGTGAGCTGACCGAGACCAACCGAGAG 2088  
DB |||||  
QY 2041 TACGTGACCGACCGGGCGCGGAGAGCTGTGAGCTGACCGAGACCAACCGAGAG 2100  
DB |||||  
QY 2089 ACCGAGCTGCAGGCCATCCAGCTGGCGCTTGCAGGACAGCGGCGAGGAGTGAACATCGTG 2148  
DB |||||  
QY 2101 ACCGAGCTGCAGGCCATCCAGCTGGCGCTTGCAGGACAGCGGCGAGGAGTGAACATCGTG 2160  
DB |||||  
QY 2149 ACCGAGCAGGTACGCGCTTGGCGCTATCCAGGCCAGCCCGACAGAGCGAGCGAG 2208  
DB |||||  
QY 2161 ACCGAGCAGGTACGCGCTTGGCGCTATCCAGGCCAGCCCGACAGAGCGAGAGCGAG 2220  
DB |||||  
QY 2209 CTGCTGAACAGATCTACGAGCAGCTGATCAAGAGGAGAGTGTGTACCTGAGCTGGTG 2268  
DB |||||  
QY 2221 CTGCTGAACAGATCTACGAGCAGCTGATCAAGAGGAGAGTGTGTACCTGAGCTGGTG 2280  
DB |||||  
QY 2269 CCGGCGCCACAGGCGCATCGCGCGCAACGAGCAGATCGAACGTGTGAGCAAGGCGCATC 2328  
DB |||||  
QY 2281 CCGGCGCCACAGGCGCATCGCGCGCAACGAGCAGATCGAACGTGTGAGCAAGGCGCATC 2340  
DB |||||  
QY 2329 CGCAAGGTGTCTTCTGAGCGGCTATCGGCGCATCGTGTATCTACAGTACATGGAC 2388  
DB |||||  
QY 2341 CGCAAGGTGTCTTCTGAGCGGCTATCGGCGCATCGTGTATCTACAGTACATGGAC 2400  
DB |||||  
QY 2389 GACCTGTACGTGGGCGCGCGCTTAGGATCGATTAAGAGCTTCCCGGGGTAGCACC 2448  
DB |||||  
QY 2401 GACCTGTACGTGGGCGCGCGCTTAGGATCGATTAAGAGCTTCCCGGGGTAGCACC 2460  
DB |||||  
QY 2449 GGTGAATTC 2457

```
Db      2461  GGTGAATTC 2469  |||||
RESULT 13
ADCI3234
ID      ADCI3234 standard; DNA; 3531 BP.
XX
XX
XX      ADCI3234;
XX
XX      18-DEC-2003 (first entry)
XX
XX      DNA of HIV construct GagPolmut_C SEQ ID NO 13.
XX
XX      expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prt.;
XX      Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
XX      Human immunodeficiency virus.
XX
XX      WO2003004620-A2.
XX
XX      16-JAN-2003.
XX
XX      05-JUL-2002; 2002WO-US021420.
XX
XX      05-JUL-2001; 2001US-0303132P.
XX
XX      31-AUG-2001; 2001US-0316860P.
XX
XX      16-JAN-2002; 2002US-0349871P.
XX
XX      (CHIR) CHIRON CORP.
XX
XX      (UYST-) UNIV STELLENBOSCH.
XX
XX      Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
XX      WPI; 2003-221593/21.
XX
XX      New expression cassette comprising a polynucleotide sequence encoding a
XX      polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
XX      Prt., or Rev polypeptide, useful for immunization, or generating
XX      packaging cell lines.
XX
XX      Disclosure; Fig 10; 301pp; English.
XX
XX      The invention relates to a novel expression cassette comprising a
XX      polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
XX      Int, Nef, p15RnaseH, Pol, Tat, Prt., or Rev polypeptide. The novel
XX      expression cassette can be used to treat HIV type C by gene therapy or
XX      used in the development of a vaccine. The gene delivery vector is
XX      administered intramuscularly, intramuscularly, intravenously,
XX      subcutaneously, intradermally, transdermally, intravaginally,
XX      intrarectally, orally or intravenously. The expression cassette is useful
XX      for immunisation, generating packaging cell lines and producing HIV
XX      polypeptides. This polynucleotide sequence represents the DNA of an HIV
XX      Type C related sequence of the invention.
XX
XX      Sequence 3531 BP; 802 A; 1210 C; 1096 G; 423 T; 0 U; 0 Other;
XX
XX      Query Match      97.9%; Score 2404.4; DB 9; Length 3531;
XX      Best Local Similarity 99.1%; Pred. No. 2.1e-289;
XX      Matches 2417; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
XX
XX      14  TGGCGAGGCGATGAGCCAGGCCACAGCCCAACATCCTGATGAGCGGAGCAACTTCA 73
XX      |||||
XX      1088 TGGCGAGGCGATGAGCCAGGCCACACACCGCTGATGATGAGAGAGCAACTTTAAA 1147
XX      |||||
XX      74  AGGGCCCCAAGCGCATCATCAAGTGTCTCACTGGCGCAAGAGGGGCCACATCGCCGCA 133
XX      |||||
XX      1148 AGGGCCCCAAGCGCATCATCAAGTGTCTCACTGGCGCAAGAGGGGCCACATCGCCGCA 1207
XX      |||||
XX      134  ATGCGCGCCCCCGCGCAAGAGGGGTGCTGGAATGCGGCAAGAGAGGGGCCACCAATGA 193
XX      |||||
XX      1208 ACTGCGCGCCCCCGCGCAAGAGGGGTGCTGGAATGCGGCAAGAGAGGGGCCACCAATGA 1267
XX      |||||

QY      194  AGGACTGCACCGAGCGCGCAGCCAACTTCTTCCGAGAGACCTGGCCTTCCCCAGGGCA 253
DB      1268  AGGACTGCACCGAGCGCGCAGCCAACTTCTTCCGAGAGACCTGGCCTTCCCCAGGGCA 1327
QY      254  AGGCGCGGAGTTCCCGAGGAGCAGAACCGCGCAACAGCCCCACAGCCCGGAGCTGC 313
DB      1328  AGGCGCGGAGTTCCCGAGGAGCAGAACCGCGCAACAGCCCCACAGCCCGGAGCTGC 1367
QY      314  AGGTGCGCGCGCAACAAACCCCGCAGAGGCGCGCGCGCGAGCGCAGGGACCCCTGAAT 373
DB      1368  AGGTGCGCGCGCAACAAACCCCGCAGAGGCGCGCGCGCGAGCGCAGGGACCCCTGAAT 1447
QY      374  TCCCCAGATCACCTGTGGCAGCGCCCTGTGGAGCATCAAGTGTGGCGGCGCAGATCA 433
DB      1448  TCCCCAGATCACCTGTGGCAGCGCCCTGTGGAGCATCAAGTGTGGCGGCGCAGATCA 1507
QY      434  AGGAGGCGCCTGTGTGACACCGGCGCGCAGCACCGTGTCTGGAGGAGATGAGCCTGCCG 493
DB      1508  AGGAGGCGCCTGTGTGACACCGGCGCGCAGCACCGTGTCTGGAGGAGATGAGCCTGCCG 1567
QY      494  GCAAGTGAAGCCCAAGATGATCGGCGCATCGGCGGCTTCATCAAGTGTGGCGCAGTACG 553
DB      1568  GCAAGTGAAGCCCAAGATGATCGGCGCATCGGCGGCTTCATCAAGTGTGGCGCAGTACG 1627
QY      554  ACCAGATCCTGATCAGATCTCGGCGCAAGAGGCGCATCGGCGCCTGTGTATCGGCCCA 613
DB      1628  ACCAGATCCTGATCAGATCTCGGCGCAAGAGGCGCATCGGCGCCTGTGTATCGGCCCA 1687
QY      614  CCCCCTGAACATATCGGCGCGCAACATCTGATCCAGCTGGGTGCGACCTGTGAATTC 673
DB      1688  CCCCCTGAACATATCGGCGCGCAACATCTGATCCAGCTGGGTGCGACCTGTGAATTC 1747
QY      674  CCATCAGCCCCCATCGAGACCGTGCCTGTAAGCTGAAGCCCGGCGATGAGCGGCCCAAG 733
DB      1748  CCATCAGCCCCCATCGAGACCGTGCCTGTAAGCTGAAGCCCGGCGATGAGCGGCCCAAG 1807
QY      734  TGAAGCAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCTGTTCGCGAGGAGA 793
DB      1808  TGAAGCAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCTGTTCGCGAGGAGA 1867
QY      794  TGGAGAGGAGGCGCAAGATCAACAAAGATCGGCGCGCGAGAACCCCTACAAACCCCGGT 853
DB      1868  TGGAGAGGAGGCGCAAGATCAACAAAGATCGGCGCGCGAGAACCCCTACAAACCCCGGT 1927
QY      854  TCGCCATCAAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGA 913
DB      1928  TCGCCATCAAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGA 1987
QY      914  ACAAGCGCACCCAGACTTCTGGAGGTGCGAGTGGGATCCGCCACCCCGCGCCCTGA 973
DB      1988  ACAAGCGCACCCAGACTTCTGGAGGTGCGAGTGGGATCCGCCACCCCGCGCCCTGA 2047
QY      974  AGAAGAGAGAGCGTGAACCGTGTGGAGCGTGGGCGAGCGCTTACTTCAAGCGTGGCCCTGG 1033
DB      2048  AGAAGAGAGAGCGTGAACCGTGTGGAGCGTGGGCGAGCGCTTACTTCAAGCGTGGCCCTGG 2107
QY      1034  ACAGAGCTTCGCAAGTACACCGCTTCACCATCCCGAGCATCAACAGAGAGACCCCG 1093
DB      2108  ACAGAGCTTCGCAAGTACACCGCTTCACCATCCCGAGCATCAACAGAGAGACCCCG 2167
QY      1094  GCATCCGTACAGTACAACTGCTGCCCGAGGGTGGAAAGGGCGAGCCCGCAGCTTTCC 1153
DB      2168  GCATCCGTACAGTACAACTGCTGCCCGAGGGTGGAAAGGGCGAGCCCGCAGCTTTCC 2227
QY      1154  AGAGCAGCATGACCAAGATCTCGGAGCGCTTCGCGCGCGCAACCCCGAGTGTGATCT 1213
DB      2228  AGAGCAGCATGACCAAGATCTCGGAGCGCTTCGCGCGCGCAACCCCGAGTGTGATCT 2287
QY      1214  ACCAGGCCCCCTGTACGTGGGCGAGCGACTGGAGATCGGCGCAGCAGCCCGCAGATCG 1273
DB      2288  ACCAGGCCCCCTGTACGTGGGCGAGCGACTGGAGATCGGCGCAGCAGCCCGCAGATCG 2347
QY      1274  AGGAGCTGCGCAAGCAGCTGCTGCGCTGGGGCTTTCACACCCCGCGACAGAGCACCAGA 1333
```

Db 2348 AGGAGTCGCAAGCACTGCTGCGCTGGGGTTTCAACACCCCGGACAAAGAACCCAGCA 2407  
Qy |||||  
Db 1334 AGGAGCCCCCTTCTGCTGCCATCGAGCTGCACCCCGACAAAGTGGACCGTGCAGCCCATCG 1393  
Qy |||||  
Db 2408 AGGAGCCCCCTTCTGCTGCCATCGAGCTGCACCCCGACAAAGTGGACCGTGCAGCCCATCG 2467  
Qy |||||  
Db 1394 AGCTGCCGAGAGAGAGAGTGCAGCGTGAAGAGCATCCAGAGCTGCTGGCGAAGCTGA 1453  
Qy |||||  
Db 2468 AGCTGCCGAGAGAGAGTGCAGCGTGAAGAGCATCCAGAGCTGCTGGCGAAGCTGA 2527  
Qy |||||  
Db 1454 ACTGGCCAGAGCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGCAGGTCTGTGCGCG 1513  
Qy |||||  
Db 2528 ACTGGCCAGAGCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGCAGGTCTGTGCGCG 2587  
Qy |||||  
Db 1514 GGGCAAGGCCCTTGACCGCATCTGTGCTGCCCTGACCGAGGAGCCGAGCTGGAGCTGGCG 1573  
Qy |||||  
Db 2588 GGGCAAGGCCCTTGACCGCATCTGTGCTGCCCTGACCGAGGAGCCGAGCTGGAGCTGGCG 2647  
Qy |||||  
Db 1574 AGAACCGGAGATCTCTGCGGAGCCGCTGACGGGTGTACTACACCCCGAGCAAGGACC 1633  
Qy |||||  
Db 2648 AGAACCGGAGATCTCTGCGGAGCCGCTGACGGGTGTACTACACCCCGAGCAAGGACC 2707  
Qy |||||  
Db 1634 TGTGTCGCGAGATCCAGAGAGGGCCACGACGATGGACCTTACAGATCTTACAGGAGC 1693  
Qy |||||  
Db 2708 TGTGTCGCGAGATCCAGAGAGGGCCACGACGATGGACCTTACAGATCTTACAGGAGC 2767  
Qy |||||  
Db 1694 CTTTCAAGAACCTGAAGACCGGCAGTACGCAAGTGGCAGCGCCGCCACCAACGAGC 1753  
Qy |||||  
Db 2768 CTTTCAAGAACCTGAAGACCGGCAGTACGCAAGTGGCAGCGCCGCCACCAACGAGC 2827  
Qy |||||  
Db 1754 TGAAGCAGCTGACCGAGGCGCTGCAGAAATGCCATGAGAGCATCTGTGATCTGGGCA 1813  
Qy |||||  
Db 2828 TGAAGCAGCTGACCGAGGCGCTGCAGAAATGCCATGAGAGCATCTGTGATCTGGGCA 2887  
Qy |||||  
Db 1814 AGACCCCAAGTTCGCTGCCATCCAGAGAGACCTTGGAGACCTTGGTGGACCGACT 1873  
Qy |||||  
Db 2888 AGACCCCAAGTTCGCTGCCATCCAGAGAGACCTTGGAGACCTTGGTGGACCGACT 2947  
Qy |||||  
Db 1874 ACTGGCAGCCACTCGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGC 1933  
Qy |||||  
Db 2948 ACTGGCAGCCACTCGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGC 3007  
Qy |||||  
Db 1934 TGTGTTACGAGTGGAGAGGAGCCCATCATCGCGCCGAGACCTTCTACGTGGAGCGCG 1993  
Qy |||||  
Db 3008 TGTGTTACGAGTGGAGAGGAGCCCATCATCGCGCCGAGACCTTCTACGTGGAGCGCG 3067  
Qy |||||  
Db 1994 CGCCCAACCGGAGACCAAGATCGGACGAGCGCGCTACGTGACCGAGCGGCGCGGACG 2053  
Qy |||||  
Db 3068 CGCCCAACCGGAGACCAAGATCGGACGAGCGCGCTACGTGACCGAGCGGCGCGGACG 3127  
Qy |||||  
Db 2054 AGATCGTGAAGCTTGACCGGAGACCAACCAAGAGACCGAGCTGACGAGCCATCCAGCTGG 2113  
Qy |||||  
Db 3128 AGATCGTGAAGCTTGACCGGAGACCAACCAAGAGACCGAGCTGACGAGCCATCCAGCTGG 3187  
Qy |||||  
Db 2114 CCTGCAAGACAGCGGAGGAGTGAACATGCTGACCGACAGCGAGTACGCTGGGCA 2173  
Qy |||||  
Db 3188 CCTGCAAGACAGCGGAGGAGTGAACATGCTGACCGACAGCGAGTACGCTGGGCA 3247  
Qy |||||  
Db 2174 TCATCAGGCCCGGACCAAGAGGAGAGCGAGCTGTGAACAGCATCTCGAGCAGC 2233  
Qy |||||  
Db 3248 TCATCAGGCCCGGACCAAGAGGAGAGCGAGCTGTGAACAGCATCTCGAGCAGC 3307  
Qy |||||  
Db 2234 TGATCAAGAGGAGAGGAGTGTACTGAGCTGGGTGCCCGCCCAAGAGGAGCATCGGCGCA 2293  
Qy |||||  
Db 3308 TGATCAAGAGGAGAGGAGTGTACTGAGCTGGGTGCCCGCCCAAGAGGAGCATCGGCGCA 3367  
Qy |||||  
Db 2294 ACAGCAGATCGCAAGAGCTGAGCAAGGGCATCCGCAAGGTGCTGTTCTTGAAGGCA 2353  
Qy |||||  
Db 3368 ACAGCAGATCGCAAGAGCTGAGCAAGGGCATCCGCAAGGTGCTGTTCTTGAAGGCA 3427  
Qy |||||  
Db 2354 TCGATGCGCGCATCTGATCTACAGTACATGACGACCTGTACGTGGGCGAGCGGGCGCC 2413  
Qy |||||

Db 3428 TCGATGGCGCATCGTGATCTACAGTACATGAGCACTGTACGTGGGCGAGCGCGCC 3487  
Qy 2414 CTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451  
Db 3488 CTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 3525  
RESULT 14  
ACA03548  
ID ACA03548 standard; DNA; 2457 BP.  
XX ACA03548;  
XX AC  
XX AC  
XX 22-MAY-2003 (first entry)  
XX Synthetic DNA encoding immunogenic HIV peptide #31.  
XX Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
KW gene therapy; packaging cell line; humoral immune response;  
KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
XX Synthetic.  
XX WO2003004657-A1.  
XX 16-JAN-2003.  
XX 05-JUL-2002; 2002WO-US021421.  
XX 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349728P.  
PR 16-JAN-2002; 2002US-0349793P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX (CHIR ) CHIRON CORP.  
XX Zur Megede J, Barnett SW, Lian Y;  
PI WPI; 2003-221602/21.  
DR New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
PT polypeptides, useful as immunogenic compositions or vaccines for  
PT generating humoral or cellular immune responses against HIV in a subject,  
PT especially humans.  
XX Example 1; Fig 36; 262pp; English.  
XX The invention describes a synthetic polynucleotide encoding 2 or more  
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
CC derived from different HIV subtypes. The polynucleotide is useful for  
CC immunisation, generation of packaging cell lines, or production of HIV  
CC polypeptides. The polynucleotide and its encoded proteins are useful as  
CC immunogenic compositions or vaccines for generating humoral or cellular  
CC immune responses against HIV in a subject, or for inducing neutralising  
CC antibodies against HIV. The gene delivery vector comprising the  
CC polynucleotide is also useful for DNA immunisation of, or for generating  
CC an immune response (e.g. a humoral or cellular immune response) in a  
CC subject such as a mammal, particularly a human. This sequence encodes a  
CC human immunodeficiency virus immunogenic peptide  
XX Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;  
SQ Query Match 97.8%; Score 2403.4; DB 7; Length 2457;  
Best Local Similarity 99.3%; Pred. No. 2.9e-289;  
Matches 2439; Conservative 0; Mismatches 6; Indels 12; Gaps 2;  
Qy 7 GCCACCATGGCGGAGGCGCATGAGCCAGCCAGCCACCATCTCTGATGACGCGAGC 66  
Db 1 GCCACCATGGCGGAGGCGCATGAGCCAGCCAGCCACCATCTCTGATGACGCGAGC 60  
Qy 67 AACTTCAAGGCCCGCCAGCGCATCATCAAGTGTCTTCACTGCGGCAAGGAGGGCCATC 126  
|||||

Db 61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGGCCACATC 120  
Qy 127 CCCCCCAACTCCCGCGCCCCCGCAAGAAAGGGTCTGTGAAGTGCGGCAAGAGGGGCCAC 186  
Db 121 GCGCGCAACTGCGCGCCCCCGCAAGAAAGGCTGTGTGAAGTGCGGCAAGAGGGGCCAC 180  
Qy 187 CAGATGAAGGACTGACCGAGCGCGCAGGCCCAACTTCTTCGCGAGGACCTGGCTTCCCC 246  
Db 181 CAGATGAAGGACTGACCGAGCGCGCAGGCCCAACTTCTTCGCGAGGACCTGGCTTCCCC 240  
Qy 247 CAGGCAAGGCGCGCGAGTTCCTCCAGAGGAGCAGAAACCGCGCAACAGCCCCACAGCGGC 306  
Db 241 CAGGCAAGGCGCGCGAGTTCCTCCAGAGGAGCAGAAACCGCGCAACAGCCCCACAGCGGC 300  
Qy 307 GAGTGTGAGGTGGCGGGCGCAAAACCCCGCAGCAGAGCGCGCGCGCGAGCGCGCAGGGCAAC 366  
Db 301 GAGTGTGAGGTGGCGGGCGCAAAACCCCGCAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCG 360  
Qy 367 CTGAACCTTCCCCCAGATCACCTGTGTGCGAGCGCGCCCCCTGTGTGAGCATCAAGGTGGCGGC 426  
Db 361 CTGAACCTTCCCCCAGATCACCTGTGTGCGAGCGCGCCCCCTGTGTGAGCATCAAGGTGGCGGC 420  
Qy 427 CAGATCAAGAGGCGCTCTGTGAACACCGCGCGCGCAGACACCGTGTGTGAGGAGATGAGC 486  
Db 421 CAGATCAAGAGGCGCTCTGTGAACACCGCGCGCGCAGACACCGTGTGTGAGGAGATGAGC 480  
Qy 487 CTGCCCCCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTCATCAAGGTGGCGGC 546  
Db 481 CTGCCCCCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTCATCAAGGTGGCGGC 540  
Qy 547 CAGTACGACCAAGTCTGATCGAGATCTGCGCGCAAGAGGCCATCGGCACCGTCTGATC 606  
Db 541 CAGTACGACCAAGTCTGATCGAGATCTGCGCGCAAGAGGCCATCGGCACCGTCTGATC 600  
Qy 607 GCGCCCGCCCGTGAACATCATCGCGCGCAACATGCTGACCGAGCTGGGTGGCGACCGCTG 666  
Db 601 GCGCCCGCCCGTGAACATCATCGCGCGCAACATGCTGACCGAGCTGGGTGGCGCTGACCGCTG 660  
Qy 667 AACTTCCCCCATCAGCCCCCATCGAGACCGTGCCTGTGAAGTCTGAAAGCCCGCGCATGAGCGGC 726  
Db 661 AACTTCCCCCATCAGCCCCCATCGAGACCGTGCCTGTGAAGTCTGAAAGCCCGCGCATGAGCGGC 720  
Qy 727 CCCAAGTGTGAGCAGTGGCCCCGTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGC 786  
Db 721 CCCAAGTGTGAGCAGTGGCCCCGTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGC 780  
Qy 787 GAGGAGTGGAGAGGAGGCGCAAGATCACCAAGATCGGCGCGCGAGAACCCCTACAAACAC 846  
Db 781 GAGGAGTGGAGAGGAGGCGCAAGATCACCAAGATCGGCGCGCGAGAACCCCTACAAACAC 840  
Qy 847 CCGCTGTTGCGCATCAAGAGAGGAGAGCAGCAGCAGAGTGGCGCAGCTGGTGAGCTTCGCG 906  
Db 841 CCGCTGTTGCGCATCAAGAGAGGAGAGCAGCAGCAGAGTGGCGCAGCTGGTGAGCTTCGCG 900  
Qy 907 GAGCTGAACAAGCGCACCGAGCTTCTGGGAGGTGACGTGGGCATCCCCCGACCCCGCGC 966  
Db 901 GAGCTGAACAAGCGCACCGAGCTTCTGGGAGGTGACGTGGGCATCCCCCGACCCCGCGC 960  
Qy 967 GGCCTGAAGAAAGAGAGCGGTGACCGTGTGTGAGCGTGGCGAGCGCCTACTTCAGCGTG 1026  
Db 961 GGCCTGAAGAAAGAGAGCGGTGACCGTGTGTGAGCGTGGCGAGCGCCTACTTCAGCGTG 1020  
Qy 1027 CCGCTGGAGCGAGCTTCGCAAGTACACCGCTTACCATCCCCCAGCATCAACAAACGAG 1086  
Db 1021 CCGCTGGAGCGAGCTTCGCAAGTACACCGCTTACCATCCCCCAGCATCAACAAACGAG 1080  
Qy 1087 ACCCGCGCATCCGTACAGTACAAAGTGTGCCCCAGGGCTGGAAGGGCAGGCCCGAC 1146  
Db 1081 ACCCGCGCATCCGTACAGTACAAAGTGTGCCCCAGGGCTGGAAGGGCAGGCCCGAC 1140  
Qy 1147 ATCTTCCAGAGCAGATGACCAAGATCTCTGAGACCTTCCGCGCGCGCAACCCCGAGATC 1206  
Db 1141 ATCTTCCAGAGCAGATGACCAAGATCTCTGAGACCTTCCGCGCGCGCAACCCCGAGATC 1200

Qy 1207 GTGATCTTACCA-----GGCCCCCTGTATCCTGGGCAAGCGACTGTGAGATCGGCCAGCAC 1260  
Db 1201 GTGATCTTACCAAGTACATGAGCGACCTGTACGTGGGCAAGCGACTGTGAGATCGGCCAGCAC 1260  
Qy 1261 CCGCGCAAGATCGAGAGCTGCGAAGCACTGTGTGCGTGGGCTTCAACCAACCCCGAC 1320  
Db 1261 CCGCGCAAGATCGAGAGCTGCGAAGCACTGTGTGCGTGGGCTTCAACCAACCCCGAC 1320  
Qy 1321 AAGAGCAACCAAGAGAGCGCCCCCTTCTGCCCCAT-----CGAGCTGCAACCCGACAG 1374  
Db 1321 AAGAGCAACCAAGAGAGCGCCCCCTTCTGTTGGATGGGCTACGAGCTGCAACCCGACAG 1380  
Qy 1375 TCGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGAACCGTGAACGACATCCAG 1434  
Db 1381 TCGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGAACCGTGAACGACATCCAG 1440  
Qy 1435 AAGCTGTGGGCAAGCTGAACTGGGCCCAAGCATCTACCCCGGATCAAGGTGGCGCAG 1494  
Db 1441 AAGCTGTGGGCAAGCTGAACTGGGCCCAAGCATCTACCCCGGATCAAGGTGGCGCAG 1500  
Qy 1495 CTGTGCAAGCTGCTGGCGGCGCAAGCCCTGACCGACATCGTCCCTTGACCGAGGAG 1554  
Db 1501 CTGTGCAAGCTGCTGGCGGCGCAAGCCCTGACCGACATCGTCCCTTGACCGAGGAG 1560  
Qy 1555 GCGAGCTGGAAGCTGGCGCGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCACCGCTGTAC 1614  
Db 1561 GCGAGCTGGAAGCTGGCGCGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCACCGCTGTAC 1620  
Qy 1615 TAGACCCCGACAGGACCTGTGTGCGCGAGATCTCAAGAGCGGCAAGTACGCCAAGATGCGC 1674  
Db 1621 TAGACCCCGACAGGACCTGTGTGCGCGAGATCTCAAGAGCGGCAAGTACGCCAAGATGCGC 1680  
Qy 1675 TACCAGATCTACAGAGCGCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAGATGCGC 1734  
Db 1681 TACCAGATCTACAGAGCGCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAGATGCGC 1740  
Qy 1735 ACCGCCCAACCAACGACGTGAAGCAGCTGACCGAGCGCGCTGCAGAAATCGCCTGGAG 1794  
Db 1741 ACCGCCCAACCAACGACGTGAAGCAGCTGACCGAGCGCGCTGCAGAAATCGCCTGGAG 1800  
Qy 1795 AGCATCGTCTATCTGGGGCAAGACCCCGCTTCCGCTGCCCATCCAGAAGAGACCTGG 1854  
Db 1801 AGCATCGTCTATCTGGGGCAAGACCCCGCTTCCGCTGCCCATCCAGAAGAGACCTGG 1860  
Qy 1855 GAGACCTGTGGAACCGACTACTTGGAGCGCCACTCGATGCCAGTGGAGTTCGTGAC 1914  
Db 1861 GAGACCTGTGGAACCGACTACTTGGAGCGCCACTCGATGCCAGTGGAGTTCGTGAC 1920  
Qy 1915 ACCCCCCCTGTGTGAAGCTGTGTACCAAGCTGGAGAGAGGCCCATCATCGCGCCGAG 1974  
Db 1921 ACCCCCCCTGTGTGAAGCTGTGTACCAAGCTGGAGAGAGGCCCATCATCGCGCCGAG 1980  
Qy 1975 ACCTTCTACGTGGAACCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTTG 2034  
Db 1981 ACCTTCTACGTGGAACCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTTG 2040  
Qy 2035 ACCGACCGGGCGCGCAGAGATCTGTAGCCTGACCGAGACCAACCAACAGAGACCGAG 2094  
Db 2041 ACCGACCGGGCGCGCAGAGATCTGTAGCCTGACCGAGACCAACCAACAGAGACCGAG 2100  
Qy 2095 CTGACGCGCATCCAGCTGGCCCTTGCAGGACAGCGGCGAGGTGAACATCGTGAACCGAC 2154  
Db 2101 CTGACGCGCATCCAGCTGGCCCTTGCAGGACAGCGGCGAGGTGAACATCGTGAACCGAC 2160  
Qy 2155 AGCCAGTACGCTTGGGCTATCTCAGGCGCGCGCAGAGGAGGAGCGGCTGGT 2214  
Db 2161 AGCCAGTACGCTTGGGCTATCTCAGGCGCGCGCAGAGGAGGAGGAGGCTGGT 2220  
Qy 2215 AACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTGTAGCTGGGTGCCCGCC 2274  
Db 2221 AACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTGTAGCTGGGTGCCCGCC 2280

QY 2275 CACAAGGGGCTCGCGGCAACGAGCGAGATCGACAAGCTGGTGAGCAAGGCGCATCCGCAAG 2334  
 Db CACAAGGGGCTCGCGGCAACGAGCGAGATCGACAAGCTGGTGAGCAAGGCGCATCCGCAAG 2340  
 QY 2335 GTGCTGTTCTTGACGCGCATCGATGGCGGCGATCGTATCTTACAGTACATGGACGACCTG 2394  
 Db GTGCTGTTCTTGACGCGCATCGATGGCGGCGATCGTATCTTACAGTACATGGACGACCTG 2400  
 QY 2395 TACGTGGGAGCGCGCCCTAGGATCGATTTAAAGCTTCCCGGGGCTAGCACCGGT 2451  
 Db TACGTGGGAGCGCGCCCTAGGATCGATTTAAAGCTTCCCGGGGCTAGCACCGGT 2457

RESULT 15

ADCL3266  
 ID ADCL3266 standard; DNA; 2457 BP.

XX AC ADCL3266;

XX 18-DEC-2003 (first entry)

XX DNA of HIV construct p2Pol-opt\_C SEQ ID NO 45.

XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
 KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.

XX Human immunodeficiency virus.

XX WO2003004620-A2.

XX 16-JAN-2003.

XX 05-JUL-2002; 2002WO-US021420.

XX 05-JUL-2001; 2001US-0303192P.

XX 31-AUG-2001; 2001US-0316860P.

XX 16-JAN-2002; 2002US-0349871P.

XX (CHIR ) CHIRON CORP.

XX (UYST-) UNIV STELLENBOSCH.

XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;

XX WPI; 2003-221593/21.

XX New expression cassette comprising a polynucleotide sequence encoding a  
 PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
 PT Prot or Rev polypeptide, useful for immunization, or generating  
 PT packaging cell lines.

XX Disclosure; Fig 42; 301pp; English.

XX The invention relates to a novel expression cassette comprising a  
 CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
 CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
 CC expression cassette can be used to treat HIV type C by gene therapy or  
 CC used in the development of a vaccine. The gene delivery vector is  
 CC administered intramuscularly, intravenously, intranasally,  
 CC subcutaneously, intradermally, transdermally, intravaginally,  
 CC intrarectally, orally or intravenously. The expression cassette is useful  
 CC for immunisation, generating packaging cell lines and producing HIV  
 CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
 CC Type C related sequence of the invention.

XX SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

Query Match 97.8%; Score 2403.4; DB 9; Length 2457;

Best Local Similarity 99.3%; Pred. No. 2.9e-289;

Matches 2439; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

QY 7 GCACACATGCCGAGGCGCATGAGCGCCAGCGCCACCAATCTGTATGAGCGCAGC 66

QY 67 AACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGAGGGGCGCATC 126  
 Db 61 AACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGAGGGGCGCATC 120  
 QY 127 GCCCGCAACTGCGCGGCCCCCGCAAGAAGGGTGTCTGGAAGTCCGCAAGAGGGGCGCAC 186  
 Db 121 GCCCGCAACTGCGCGGCCCCCGCAAGAAGGGTGTCTGGAAGTCCGCAAGAGGGGCGCAC 180  
 QY 187 CAGATGAAGGACTGCACCGAGCGCGCAGGCCAACTTCTTCCGCGAGGACCTTGGCCTTCCCC 246  
 Db 181 CAGATGAAGGACTGCACCGAGCGCGCAGGCCAACTTCTTCCGCGAGGACCTTGGCCTTCCCC 240  
 QY 247 CAGGCAAGGCCCCCGGAGTTTCCCCAGCGAGCAGAAACCGCGCCAAACAGCCCCACAGCGGC 306  
 Db 241 CAGGCAAGGCCCCCGGAGTTTCCCCAGCGAGCAGAAACCGCGCCAAACAGCCCCACAGCGGC 300  
 QY 307 GAGCTGCAGGTGCGCGGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCAC 366  
 Db 301 GAGCTGCAGGTGCGCGGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCAC 360  
 QY 367 CTGAATTTCCCCAGATCACCCCTGTGGACGCGCCCCCTGTGTGAGCATCAAGGTGGCGGC 426  
 Db 361 CTGAATTTCCCCAGATCACCCCTGTGGACGCGCCCCCTGTGTGAGCATCAAGGTGGCGGC 420  
 QY 427 CAGATCAAGGAGGCGCTGTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 486  
 Db 421 CAGATCAAGGAGGCGCTGTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 480  
 QY 487 CTGCGCGCAAGTGGAAAGCCCAAGATGATCGCGGCGCATCGCGGCTTTCATCAAGTGC 546  
 Db 481 CTGCGCGCAAGTGGAAAGCCCAAGATGATCGCGGCGCATCGCGGCTTTCATCAAGTGC 540  
 QY 547 CAGTACGACGAGATCTTGATCGAGATCTGCGGCAAGAGGCGCATCGCGCGCTGTGATC 606  
 Db 541 CAGTACGACGAGATCTTGATCGAGATCTGCGGCAAGAGGCGCATCGCGCGCTGTGATC 600  
 QY 607 GCGCCCAACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCG 666  
 Db 601 GCGCCCAACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCG 660  
 QY 667 AACTTCCCATCAGCGCGCATCGAGACCGTGCCTGCGCGCGCATCGAGCGCGCGCGCGC 726  
 Db 661 AACTTCCCATCAGCGCGCATCGAGACCGTGCCTGCGCGCGCATCGAGCGCGCGCGCGC 720  
 QY 727 CCCAAGGTGAAGCAGTGGCGGCTTCCAGCGAGGAGAGATCAAGGCGCTGACCGCGCATCTGC 786  
 Db 721 CCCAAGGTGAAGCAGTGGCGGCTTCCAGCGAGGAGAGATCAAGGCGCTGACCGCGCATCTGC 780  
 QY 787 GAGGAGATGGAGAGGAGGCGCAAGATCAACAGATCGCGCGCGCGCGCGCGCGCGCGCAC 846  
 Db 781 GAGGAGATGGAGAGGAGGCGCAAGATCAACAGATCGCGCGCGCGCGCGCGCGCGCGCAC 840  
 QY 847 CCGGTGTTCCGCATCAAGAGAGAGAGCAGCACCAAGTGGCGCGCGCGCGCGCGCGCGCGC 906  
 Db 841 CCGGTGTTCCGCATCAAGAGAGAGAGCAGCACCAAGTGGCGCGCGCGCGCGCGCGCGCGC 900  
 QY 907 GAGCTGAACAGCGCACCCAGGACTTCTGGAGAGTGCAGCTGGGCGCATCCCCCGCGCGC 966  
 Db 901 GAGCTGAACAGCGCACCCAGGACTTCTGGAGAGTGCAGCTGGGCGCATCCCCCGCGCGC 960  
 QY 967 GCGCTGAAGAGAGAGAGGCGTACCGTGTCTGACGCTGGCGCGCGCGCGCGCGCGCGCG 1026  
 Db 961 GCGCTGAAGAGAGAGAGGCGTACCGTGTCTGACGCTGGCGCGCGCGCGCGCGCGCGCG 1020  
 QY 1027 CCGCTGAAGAGAGGACTTCCCGCAAGTACACCGCGCTTCCACCATCCCGCGCATCAACAGCGAG 1086  
 Db 1021 CCGCTGAAGAGGAGACTTCCCGCAAGTACACCGCGCTTCCACCATCCCGCGCATCAACAGCGAG 1080  
 QY 1087 ACCCGCGGCTCCGCTACAGTACAAGTGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1146



QY 1147 ATCTCCAGCAGATGACCAAGATCTTGGAGCCCTTCGCGCCGCAACCCCGAGATC 1206  
 Db 1141 ATCTCCAGCAGATGACCAAGATCTTGGAGCCCTTCGCGCCGCAACCCCGAGATC 1200  
 QY 1207 GTGATCTACCA-----GCGCCCTCTGTACGTGGGCGAGCAGCTGGAGATCGGCCAGCAC 1260  
 Db 1201 GTGATCTACCAAGTACATGACGACCTGTACGTGGGCGAGCAGCTGGAGATCGGCCAGCAC 1260  
 QY 1261 CCGGCCAAGATCGAGAGTGGCAAGCACTGTCTGGGTGGGCTTACCAACCCCGAC 1320  
 Db 1261 CCGGCCAAGATCGAGAGTGGCAAGCACTGTCTGGGTGGGCTTACCAACCCCGAC 1320  
 QY 1321 AAGAAGCACAAGAGAGCCCTCTCTGCCCCAT-----CGAGCTGCACCCCGACAAG 1374  
 Db 1321 AAGAAGCACAAGAGAGCCCTCTCTGCCCCAT-----CGAGCTGCACCCCGACAAG 1380  
 QY 1375 TGGACGTGAGCCCATCGAGCTGCCGAGAGAGAGTGGACCGTGAAGCAATCCAG 1434  
 Db 1381 TGGACGTGAGCCCATCGAGCTGCCGAGAGAGAGTGGACCGTGAAGCAATCCAG 1440  
 QY 1435 AAGCTGCTGGCAAGCTGAATGGGCGAGCCAGATCTACCCGCGCATCAAGTGGCGCAG 1494  
 Db 1441 AAGCTGCTGGCAAGCTGAATGGGCGAGCCAGATCTACCCGCGCATCAAGTGGCGCAG 1500  
 QY 1495 CTGTGCAAGCTGCTGGCGGCGCAAGGCCCTTGAACGACATGCTGCCCTGACCCGAGAG 1554  
 Db 1501 CTGTGCAAGCTGCTGGCGGCGCAAGGCCCTTGAACGACATGCTGCCCTGACCCGAGAG 1560  
 QY 1555 GCGGAGCTGAGCTGGCGAGAACCGGAGATCTTGGCGGCGCCGCTGACGGGCTGTAC 1614  
 Db 1561 GCGGAGCTGAGCTGGCGAGAACCGGAGATCTTGGCGGCGCCGCTGACGGGCTGTAC 1620  
 QY 1615 TACGACCCCGAAGGACTGTGTGCGGAGATCCAGAGCAGGGCCACGACCAAGTGAC 1674  
 Db 1621 TACGACCCCGAAGGACTGTGTGCGGAGATCCAGAGCAGGGCCACGACCAAGTGAC 1680  
 QY 1675 TACGAGATCTACGAGGCGCTTCAAGACCTGAGACCGGCGCAGTACGCCAAGATGCGC 1734  
 Db 1681 TACGAGATCTACGAGGCGCTTCAAGACCTGAGACCGGCGCAGTACGCCAAGATGCGC 1740  
 QY 1735 ACCGCCACACCAAGCAGTGAAGCAGTGTGACCGAGGCGCTGAGAGATGCCATGGAG 1794  
 Db 1741 ACCGCCACACCAAGCAGTGAAGCAGTGTGACCGAGGCGCTGAGAGATGCCATGGAG 1800  
 QY 1795 AGCATGTGTATCTGGGCGAGACCCCGCAAGTTCGCCCTGCCATCCAGAGGAGCCTGG 1854  
 Db 1801 AGCATGTGTATCTGGGCGAGACCCCGCAAGTTCGCCCTGCCATCCAGAGGAGCCTGG 1860  
 QY 1855 GAGACTGTGACCGGACTACTTGGCAGGCGCCACTGGATCCCGAGTGGGAGTTGCTGAAC 1914  
 Db 1861 GAGACTGTGACCGGACTACTTGGCAGGCGCCACTGGATCCCGAGTGGGAGTTGCTGAAC 1920  
 QY 1915 ACCGCCCTCTGTGAGTGTGTACAGTGTGAGAGAGGCCCATCATCGGCGCGAG 1974  
 Db 1921 ACCGCCCTCTGTGAGTGTGTACAGTGTGAGAGAGGCCCATCATCGGCGCGAG 1980  
 QY 1975 ACTTCTTACGTGAGCGGCGCCCAACCGGAGACCAAGATCGGCAAGCGGCTACGTG 2034  
 Db 1981 ACTTCTTACGTGAGCGGCGCCCAACCGGAGACCAAGATCGGCAAGCGGCTACGTG 2040  
 QY 2035 ACCGACCGGGCGCGGAGAGATCTGTGACCTGACCGAGACCAACCAAGAGACCGAG 2094  
 Db 2041 ACCGACCGGGCGCGGAGAGATCTGTGACCTGACCGAGACCAACCAAGAGACCGAG 2100  
 QY 2095 CTGACAGGCGCATCCAGTGGCTGAGGACAGCGGCGAGGTTGAACATGCTGACCGAC 2154  
 Db 2101 CTGACAGGCGCATCCAGTGGCTGAGGACAGCGGCGAGGTTGAACATGCTGACCGAC 2160  
 QY 2155 AGCCAGTACGCCCTGGGATCATCCAGGCCAGCCCGCAAGAGCGGAGCGAGCTGGTG 2214  
 Db 2161 AGCCAGTACGCCCTGGGATCATCCAGGCCAGCCCGCAAGAGCGGAGCGAGCTGGTG 2220  
 QY 2215 AACCAGATCATCGAGCAGCTGATCAAGAGGAGAGTGTACTCTGAGCTGGTGGCCGCC 2274

Db 2221 AACAGATCATCGAGCAGCTGTATCAAGAGGAGAGGTGTACTGAGCTGGGTGCCCGCC 2280  
 QY 2275 CACAAGGGCATCGCGCGCAACCGAGCAGATCGACAGCTGTGAGCAAGGGCATCGGCAAG 2334  
 Db 2281 CACAAGGGCATCGCGCGCAACCGAGCAGATCGACAGCTGTGAGCAAGGGCATCGGCAAG 2340  
 QY 2335 GTGCTGTTCTTGGACCGGATCGATGGCGGATCGATGATCTACAGTACATGAGCAGCCTG 2394  
 Db 2341 GTGCTGTTCTTGGACCGGATCGATGGCGGATCGATGATCTACAGTACATGAGCAGCCTG 2400  
 QY 2395 TACGTGGGCGAGCGCGCGCCCTAGGATCGATTAAGAGCTTCCCGGGGCTAGCACCCTG 2451  
 Db 2401 TACGTGGGCGAGCGCGCGCCCTAGGATCGATTAAGAGCTTCCCGGGGCTAGCACCCTG 2457

Search completed: April 10, 2004, 07:32:35  
 Job time : 626.487 secs